

## ORBIT - Online Repository of Birkbeck Institutional Theses

---

Enabling Open Access to Birkbeck's Research Degree output

### Negations without "not" : alternative forms of negation and contrast classes in conditional inference

<https://eprints.bbk.ac.uk/id/eprint/40321/>

Version: Full Version

**Citation: Vance, James R. H. (2018) Negations without "not" : alternative forms of negation and contrast classes in conditional inference. [Thesis] (Unpublished)**

© 2020 The Author(s)

---

All material available through ORBIT is protected by intellectual property law, including copyright law.

Any use made of the contents should comply with the relevant law.

---

Negations Without “Not”:  
Alternative forms of Negation and Contrast Classes  
in Conditional Inference

James Richard Hazlett Vance

Doctor of Philosophy  
Birkbeck, University of London  
2018

## **Declaration**

I confirm that the work submitted in this thesis is my own and I confirm that the thesis references information derived from other sources.

James Vance

## Acknowledgements

I would like to thank my supervisor, Mike Oaksford, first for getting me interested in reasoning and then having providing support for about a decade through BSc, MRes and finally guiding me through this PhD.

I would like to thank the other people that have made working in Birkbeck particularly fun – especially Alex Shepherd and the teaching assistants I have worked with on the critical analysis course (which provides a useful set of tools to consider and critique my work).

I am also grateful for the support of the managers, colleagues, and teams that I have worked with over the past few years at the Department for Education, the Education and Skills Funding Agency and LocatED.

I'd like to thank my friends from outside work and study who have supported and helpfully distracted me at various points from both of those activities, particularly Alex Warren and Neil Brown.

I would like to thank my mother, Maggie Vance, and my father, Andrew Vance, for their support and encouragement. Finally, I'd like to thank Gillian Machin for her love and tolerance, particularly as I've spent late nights finishing this thesis.



## **Abstract**

This thesis explored the role of negation in conditional reasoning by looking at how different types of negation affected responses to reasoning questions. Previous research has observed several systematic biases in conditional reasoning with negations (Manktelow, 2012). Prior reasoning research has looked at negation as a simple logical operator. However, negation can take a range of forms and provide different implications (Horn, 1989). These experiments tested the effect of different types of negation on conditional reasoning.

The first set of experiments looked at how using different types of negation in propositions affected people's probability ratings of those propositions. The three experiments identified differences when using several different types of negation. However, the differences observed were small. These results confirmed that people do take different implications from different types of negation.

The second set of four experiments used an extension of Evans' (1977) conditional inference paradigm to compare two types of negation (using "not" and affix "un-"). The experiments looked at whether responses reflected the different probabilistic implications of those negative forms. Any effect of different types of negation appears overwhelmed by plausibility and other material effects. Broadly these results provide are consistent with probabilistic models of reasoning. However, inconsistencies in the results suggest further work is necessary to rule out other models.

The final set of three experiments used a novel learning task to test the effect of frequency information on inference endorsement tasks. The first experiment confirmed Oaksford and Chater's (2007) prediction of MP inference suppression when implicit negation is used instead of "not" negation and participants are provided with appropriate frequency information. Two further experiments confirmed that this effect was the result of the frequency information and extended it to AC, DA and MT inferences. These results provide support for a probabilistic approach to reasoning.

# Contents

Declaration .....	2
Acknowledgements .....	3
Abstract .....	4
Contents .....	5
List of Tables.....	12
List of Figures .....	14
1. Conditional Reasoning Biases and Negation .....	16
1.1. Conditional Reasoning .....	17
1.2. Systematic Biases in Conditional Reasoning .....	19
1.2.1. Biases Involving Negation .....	20
1.2.1.1. Matching Bias .....	20
1.2.1.2. Negative Conclusion Bias .....	22
1.2.2. Conditional or Biconditional? .....	24
1.3. Accounts of Reasoning.....	25
1.3.1. Mental Logics.....	26
1.3.2. Mental Models .....	29
1.3.3. Probabilistic Approach.....	32
1.3.4. Multiple Processes .....	34
1.3.5. Alternative Models.....	36
1.4. Accounts of Negation .....	37
1.4.1. Logical Operator .....	37
1.4.1.1. Other Logics.....	38
1.4.2. Natural Language .....	40
1.4.2.1. Scope of Negation .....	41
1.4.2.2. Pragmatic Ambiguity .....	42
1.4.3. Contrast Classes .....	43
1.5. The Psychology of Negation .....	44
1.5.1. Processing Time .....	44
1.5.1.1. Contexts of Plausible Denial.....	44
1.5.2. Probing Negations.....	45
1.6. Approach .....	46

2.	Different Types of Negation and Probability.....	48
2.1.	Reasoning and Probability.....	48
2.1.1.	Probability Rating Tasks.....	52
2.2.	Different Types of Negation.....	53
2.2.1.	Affixal Negation.....	54
2.2.2.	Implicit Negation .....	57
2.3.	Experiment 1 .....	60
2.3.1.	Predictions.....	63
2.3.2.	Method .....	63
2.3.2.1.	Design .....	63
2.3.2.2.	Participants.....	64
2.3.2.3.	Materials.....	64
2.3.2.4.	Procedure.....	65
2.3.3.	Results .....	66
2.3.3.1.	Overall Differences between Different Types of Negation .....	66
2.3.3.2.	Affixal Negation.....	67
2.3.3.3.	Implicit Negation .....	68
2.3.4.	Discussion .....	69
2.4.	Experiment 2 .....	71
2.4.1.1.	Online Recruitment of Participants.....	73
2.4.2.	Predictions.....	73
2.4.3.	Method .....	74
2.4.3.1.	Design .....	74
2.4.3.2.	Participants.....	74
2.4.3.3.	Materials.....	74
2.4.3.4.	Procedure.....	75
2.4.4.	Results .....	76
2.4.4.1.	Responses to Each Scenario.....	77
2.4.5.	Discussion .....	79
2.5.	Experiment 3 .....	81
2.5.1.	Predictions.....	83
2.5.2.	Method .....	84
2.5.2.1.	Design .....	84
2.5.2.2.	Participants.....	84

2.5.2.3. Materials.....	85
2.5.2.4. Procedure.....	87
2.5.3. Results.....	87
2.5.3.1. Probability Rating Tasks – Overall Responses .....	87
2.5.3.2. Probability Rating Tasks – Responses to Each Scenario .....	89
2.5.3.3. Probability Rating Tasks – Possible Other Factors.....	89
2.5.3.4. Plausibility Ratings of Conditional Rules.....	91
2.5.4. Discussion .....	92
2.6. General Discussion.....	94
3. Different Types of Negation in Conditional Inference .....	97
3.1. Evans’ (1977) Negation Paradigm .....	97
3.1.1. Extending the Negation Paradigm .....	100
3.2. Causal Relationships and Plausibility .....	101
3.2.1. The Conditional as Causal Relationship .....	102
3.2.2. Context and Plausibility in Causal Conditionals.....	103
3.2.3. Causal Models.....	105
3.3. Experiment 1 .....	110
3.3.1. Predictions.....	111
3.3.2. Method .....	114
3.3.2.1. Design .....	114
3.3.2.2. Participants.....	114
3.3.2.3. Materials.....	114
3.3.2.4. Procedure.....	116
3.3.3. Results.....	117
3.3.3.1. Conditional Inference – Levels of Endorsement.....	117
3.3.3.2. Effect of Polarity on Plausibility Ratings.....	119
3.3.3.3. Interaction Between Plausibility and Conditional Inference .....	122
3.3.4. Discussion .....	123
3.4. Experiment 2 .....	126
3.4.1. Predictions.....	127
3.4.2. Method .....	128
3.4.2.1. Design .....	128
3.4.2.2. Participants.....	128
3.4.2.3. Materials.....	129

3.4.2.4. Procedure.....	131
3.4.3. Results.....	132
3.4.3.1. Results for Each Scenario .....	134
3.4.4. Discussion .....	138
3.5. Experiment 3 .....	141
3.5.1. Predictions.....	142
3.5.2. Method .....	143
3.5.2.1. Design .....	143
3.5.2.2. Participants.....	143
3.5.2.3. Materials.....	143
3.5.2.4. Procedure.....	145
3.5.3. Results .....	146
3.5.3.1. Pattern of Inference Scores .....	146
3.5.3.2. Negative Conclusion Bias .....	150
3.5.3.3. Role of Plausibility.....	150
3.5.3.4. Other Possible Factors .....	154
3.5.4. Discussion .....	156
3.6. Experiment 4 .....	159
3.6.1. Predictions.....	160
3.6.2. Method .....	161
3.6.2.1. Design .....	161
3.6.2.2. Participants.....	161
3.6.2.3. Materials.....	162
3.6.2.4. Procedure.....	162
3.6.3. Results .....	163
3.6.3.1. Pattern of Inference Scores .....	164
3.6.3.2. Negative Conclusion Bias .....	167
3.6.3.3. Role of Plausibility.....	169
3.6.3.4. Other Possible Factors .....	171
3.6.4. Discussion .....	173
3.7. General Discussion.....	176
4. Contrast Classes in Conditional Inference .....	179
4.1. The Contrast Class Account of Negation .....	179
4.1.1. Contrast Classes in Reasoning .....	181

4.1.2.	Implicit Negations Using Contrast Class Members .....	187
4.1.3.	Implicit negation effects and other theories of reasoning .....	191
4.2.	Experiment 1 .....	193
4.2.1.	Approach to providing contextual information.....	193
4.2.2.	Predictions.....	195
4.2.3.	Method .....	196
4.2.3.1.	Design .....	196
4.2.3.2.	Participants.....	196
4.2.3.3.	Materials.....	197
4.2.3.4.	Procedure.....	198
4.2.4.	Results .....	199
4.2.4.1.	Engagement with Learning Phase.....	201
4.2.4.2.	Participants' Probability Estimates .....	204
4.2.5.	Discussion .....	205
4.3.	Experiment 2 .....	206
4.3.1.	Predictions.....	207
4.3.2.	Method .....	209
4.3.2.1.	Design .....	209
4.3.2.2.	Participants.....	210
4.3.2.3.	Materials.....	210
4.3.2.4.	Procedure.....	212
4.3.3.	Results .....	213
4.3.3.1.	Engagement with Learning Phase.....	215
4.3.3.2.	Materials Effects .....	218
4.3.4.	Discussion .....	220
4.4.	Experiment 3 .....	221
4.4.1.	Predictions.....	222
4.4.2.	Method .....	224
4.4.2.1.	Design .....	224
4.4.2.2.	Participants.....	225
4.4.2.3.	Materials.....	225
4.4.2.4.	Procedure.....	227
4.4.3.	Results .....	228
4.4.3.1.	Engagement with Learning Phase.....	230

4.4.4. Discussion .....	232
4.5. General Discussion.....	234
5. Negation in Reasoning – Conclusions and Next Steps .....	236
5.1. Key Findings .....	236
5.1.1. Different Types of Negation and Probability.....	236
5.1.2. Different Types of Negation in Conditional Inference .....	238
5.1.3. Contrast Classes in Conditional Inference .....	240
5.2. Accounts of Reasoning.....	242
5.2.1. Mental Logics.....	242
5.2.2. Mental Models .....	244
5.2.3. Probabilistic Approach.....	245
5.2.4. Multiple Processes .....	246
5.2.5. Alternative Models.....	246
5.3. Next Steps.....	247
References .....	249
List of Appendices .....	260
Appendix 1. Materials Used in Chapter 2 Experiments.....	261
A1.1. Experiment 1 .....	261
A1.1.1. Task Instructions .....	261
A1.1.2. Task Questions .....	261
A1.2. Experiment 2 .....	268
A1.2.1. Task Instructions .....	268
A1.2.2. Task Questions .....	268
A1.3. Experiment 3 .....	270
A1.3.1. Task Instructions for Part 1 .....	271
A1.3.2. Task Questions for Part 1 .....	271
A1.3.3. Task Instructions for Part 2 .....	275
A1.3.4. Task Questions for Part 2 .....	276
Appendix 2. Materials Used in Chapter 3 Experiments.....	281
A2.1. Experiment 1 .....	281
A2.1.1. Task Instructions for Part 1 .....	281
A2.1.2. Task Questions for Part 1 .....	282
A2.1.3. Task Instructions for Part 2 .....	283
A2.1.4. Task Questions for Part 2 .....	284

A2.2. Experiment 2 .....	284
A2.2.1. Task Instructions .....	285
A2.2.2. Practice Questions .....	286
A2.2.3. Task Questions .....	287
A2.3. Experiments 3 and 4 .....	289
A2.3.1. Task Instructions for Experiments 3 and 4, Part 1 .....	289
A2.3.2. Task Questions for Experiments 3 and 4, Part 1 .....	290
A2.3.3. Task Instructions for Experiments 3 and 4, Part 2 .....	306
A2.3.4. Task Questions for Experiments 3 and 4, Part 2 .....	307
Appendix 3. Materials Used in Chapter 4 Experiments .....	312
A3.1. ‘Animals’ Scenario .....	312
A3.1.1. Task Instructions for Experiment 1 .....	312
A3.1.2. Task Instructions for Experiment 2 .....	315
A3.1.3. Learning Phase .....	317
A3.1.4. Inference Questions .....	330
A3.1.5. Verification Task .....	331
A3.2. ‘Vehicles’ Scenario .....	332
A3.2.1. Task Instructions .....	332
A3.2.2. Learning Phase .....	335
A3.2.3. Inference Questions .....	345
A3.2.4. Verification Task .....	347
A3.3. ‘Shapes’ Scenario .....	347
A3.3.1. Task Instructions .....	347
A3.3.2. Learning Phase .....	350
A3.3.3. Inference Questions .....	359
A3.3.4. Verification Task .....	361



## List of Tables

Table 2.1 – Summary of responses to chapter 2, experiment 1 .....	66
Table 2.2 – Comparing proportion estimates from Bianchi et al. (2011) and present experiment.....	69
Table 2.3 – Summary of responses in chapter 2, experiment 2 .....	76
Table 2.4 – Mean responses to each scenario in chapter 2, experiment 2 .....	78
Table 2.5 – Responses to probability rating tasks in chapter 2, experiment 3 .....	88
Table 2.6 – Summary of responses to plausibility tasks in chapter 2, experiment 3 .....	91
Table 3.1 – Comparisons between antecedent and consequent polarity conditions .....	121
Table 3.2 – Effects of antecedent and consequent negation on plausibility of conditional .....	122
Table 3.3 – Varying polarity conditions .....	131
Table 3.4 – Percentage of inferences endorsed in experiment 2.....	132
Table 3.5 – Percentage of inferences endorsed using scenario 1 .....	135
Table 3.6 – Percentage of inferences endorsed using scenario 2 .....	137
Table 3.7 – Mean (standard deviation) inference scores by inference type, antecedent polarity and consequent polarity .....	147
Table 3.8 – Mean (standard deviation) responses to inference questions in chapter 3, experiment 3.....	150
Table 3.9 – Mean plausibility ratings by antecedent polarity and consequent polarity	151
Table 3.10 – Proportion of responses in each polarity condition allocated to each plausibility group .....	153
Table 3.11 – Mean (standard deviation) inference scores by inference type, antecedent polarity and consequent polarity .....	164
Table 3.12 – Mean (standard deviation) responses to inference questions in chapter 3, experiment 4.....	168
Table 3.13 – Mean plausibility ratings of conditionals by antecedent polarity and consequent polarity .....	169
Table 3.14 – Mean plausibility ratings of antecedent statements by polarity.....	169
Table 4.1 – Oaksford and Chater’s (2007) illustrative examples of contingency tables for a conditional rule where, e.g. $\neg a = \{b\ c\}$ .....	190
Table 4.2 – Contextual information provided about animals.....	195
Table 4.3 – Summary of responses to chapter 4, experiment 1 .....	200

Table 4.4 – Contextual information provided – MP manipulation .....	208
Table 4.5– Contextual information provided – AC manipulation .....	209
Table 4.6 – Summary of responses to chapter 4, experiment 2 .....	214
Table 4.7 – Contextual information provided – DA manipulation .....	223
Table 4.8 – Contextual information provided – MT manipulation.....	224
Table 4.9 – Summary of responses to chapter 4, experiment 3 .....	229
Table A1.1 – Tasks in chapter 2, experiment 1.....	262
Table A1.2 – Tasks in chapter 2, experiment 2.....	269
Table A1.3 – Tasks in chapter 2, experiment 3, part 1 .....	272
Table A1.4 – Tasks in chapter 2, experiment 3, part 2 .....	277
Table A2.1 – Scenarios in chapter 3, experiment 1, part 1 .....	282
Table A2.2 – Polarity conditions in chapter 3, experiment 1, part 1 .....	283
Table A2.3 – Polarity conditions in chapter 3, experiment 1, part 2 .....	284
Table A2.4 – Scenarios in chapter 3, experiment 2 .....	288
Table A2.5 – Polarity conditions in chapter 3, experiment 2.....	288
Table A2.6 – Scenario introductions in chapter 3, experiments 3 and 4 .....	292
Table A2.7 – Inference tasks in chapter 3, experiments 3 and 4, part 1 .....	294
Table A2.8 – Inference tasks in chapter 3, experiments 3 and 4, part 2 .....	308
Table A3.1 – Learning task stimuli for ‘animals’ scenario.....	318
Table A3.2 – Inference task stimuli for ‘animals’ scenario .....	331
Table A3.3 – Learning task stimuli for ‘vehicles’ scenario .....	336
Table A3.4 – Inference task stimuli for ‘vehicles’ scenario .....	346
Table A3.5 – Learning task stimuli for ‘shapes’ scenario .....	351
Table A3.6 – Inference task stimuli for ‘shapes’ scenario.....	360

## List of Figures

Figure 3.1 – Fernbach and Erb’s (2013) CBNs for (i) simple MP, (ii) extended MP and (iii) AC inferences.....	106
Figure 3.2 – Chart showing percentage of inferences by inference and antecedent polarity in experiment 1. ....	119
Figure 3.3 – Chart showing mean plausibility ratings by antecedent and consequent polarities in experiment 1.....	120
Figure 3.4 – Chart showing percentage of inferences by inference and antecedent polarity in experiment 2 (combined scenarios).....	133
Figure 3.5 – Chart showing interaction between inference type and antecedent polarity (scenario 1).....	136
Figure 3.6 – Chart showing interaction between inference type and consequent polarity (scenario 1).....	137
Figure 3.7 – Chart showing interaction between inference type and antecedent polarity on inference responses in experiment 3. ....	148
Figure 3.8 – Chart showing interaction between inference type and consequent polarity on inference responses in experiment 3. ....	149
Figure 3.9 – Chart showing interaction between antecedent polarity and consequent polarity on inference responses in experiment 3.....	149
Figure 3.10 – Chart showing interaction effect between antecedent and consequent polarity on probability ratings of conditionals in experiment 3.....	152
Figure 3.11 – Chart showing mean inference responses in experiment 3 by inference type and standard negation paradigm conditional formats. ....	156
Figure 3.12 – Chart showing interaction between inference type and antecedent polarity on inference responses in experiment 4. ....	166
Figure 3.13 – Chart showing interaction between inference type and consequent polarity on inference responses in experiment 4. ....	166
Figure 3.14 – Chart showing interaction between antecedent polarity and consequent polarity on inference responses in experiment 4.....	167
Figure 3.15 – Chart showing interaction effect between antecedent and consequent polarity on probability ratings of conditionals in experiment 4.....	171
Figure 3.16 – Chart showing mean inference responses in experiment 4 by inference type and standard negation paradigm conditional formats. ....	174

Figure 4.1 – Chart Showing Mean Confidence Scores for MP and AC Inferences of Participants that Demonstrated Learning in Experiment 1. ....	204
Figure 4.2 – Chart Showing Mean Confidence Scores for MP and AC Inferences of Participants that Demonstrated Learning in Experiment 2. ....	218
Figure 4.3 – Chart Showing Mean Confidence Scores for DA and MT Inferences of Participants that Demonstrated Learning in Experiment 3. ....	232

# 1. Conditional Reasoning Biases and Negation

In everyday life, people need to infer the unknown from the known. They need to consider the facts that they know and draw conclusions on which basis they can act. For example, a person may know that if they flick a light switch, a light will come on. Based on this rule, they act to turn on the switch based on the conclusion that the light will come on. As this example illustrates, potentially the most important form of reasoning uses conditionals, sentences of the form ‘if  $p$ , then  $q$ ’ (Oaksford & Chater, 2010). Conditional sentences and the inferences they licence can be defined logically. However, human reasoning often shows biases that are inconsistent with the classical logic interpretation of the conditional (Manktelow, 2012). Some have suggested that such results undermine the assumption that humans are rational (Stich, 1985) but others suggest that the biases are consistent with alternative standards of rationality (Oaksford & Chater, 1994). Understanding how people reason with conditionals is therefore central to our understanding of humans as an animal capable of rational thought, as Aristotle argued.

Several biases observed in conditional reasoning are associated with the use of negations (e.g., Evans, 1972a; Evans, 1977). Negation is a concept that has a clear meaning in classical logic but a range of different implications in everyday natural language (Horn, 1989).

The present study aims to explore the effects that the use of negation has on conditional reasoning and what they can tell us about how people reason.

In this initial chapter, I will discuss the background to the present study. Firstly, I will discuss prior research in conditional reasoning and what previous studies involving negation in conditional inferences have found. I will consider the different models that have been proposed to account for the biases observed in human reasoning. I will then describe some different accounts of negation and how prior research has explored the processing of negation. Finally, I will set out the structure of the current thesis and outline the approach that the present research will take.

## 1.1. Conditional Reasoning

Conditional sentences describe a relationship between two propositions in which one proposition must be true if the other is true. Being able to describe and use such relationships is crucial to understanding of causal structures (e.g. *If I flick the switch, the light will come on*), law-like behaviour (e.g. *If Gill buys lunch at Pret, she gets a cappuccino*), conditional promises (e.g. *If you give me a lift tonight, you can keep the car for the rest of the week*), regulations (e.g. *If you drive, you must be aged at least 17*) and alternative possibilities (e.g. *If it had not rained, Button would have won the race*) (Oaksford & Chater, 2010). People need to be able to understand and use the relationships codified in conditional sentences to draw inferences about states of affairs that go beyond the known premises. Conditional rules are expressed in different, logically equivalent, forms including ‘if  $p$  then  $q$ ’, ‘ $p$  only if  $q$ ’, ‘ $q$  if  $p$ ’ or symbolically as ‘ $p \rightarrow q$ ’. According to classical logic, material conditionals are truth functions (‘if  $p$  then  $q$ ’ is false only if  $p$  is true and  $q$  is false, otherwise it is true).

Several tasks have been used to look at how people use conditionals in reasoning, asking them to evaluate, construct or select terms in a conditional inference. In a rule of the form ‘if  $p$  then  $q$ ’, the  $p$  term is referred to as the antecedent and the  $q$  term as the consequent.

Evaluation (or verification) tasks often provide participants with an inference (which consists of: the major premise – the conditional sentence; the minor premise – a proposition related to either the antecedent or consequent; and, a conclusion – a proposition about the other term not referenced in the minor premise). They ask participants whether they would endorse the conclusion or not. For example, using abstract materials, Evans (1977) presented two statements like ‘If the letter is G, then the number is not 9’ and ‘The number is not 9’ and a conclusion like ‘The letter is G’. In each task, participants were asked to decide whether the conclusion followed from the premises. Four forms of inference have typically been investigated using conditionals: Modus Ponens (MP) and Modus Tollens (MT), which are valid forms of inference under classical logic, and invalid forms Denying the Antecedent (DA) and Affirming the Consequent (AC). These take the following logical forms (symbolically ‘ $\neg$ ’ is used to negate the following proposition and ‘ $\therefore$ ’ means ‘therefore’):

- MP:  $p \rightarrow q, p \therefore q$
- MT:  $p \rightarrow q, \neg q \therefore \neg p$
- DA:  $p \rightarrow q, \neg p \therefore \neg q$
- AC:  $p \rightarrow q, q \therefore p$

While DA and AC are not valid inferences based on conditionals, the premises do not preclude the conclusions occurring (they are simply insufficient to infer the given conclusion, although see §1.2.2 below for a discussion of the biconditional interpretation of conditional relationships). While there are other valid and invalid inferences, some combinations of conditional, additional premise and conclusion have a conclusion that is mutually exclusive of the premises (e.g.  $p \rightarrow q, p \therefore \neg q$ ) or irrelevant (e.g.  $p \rightarrow q, p \therefore r$ ).

Construction (or generation) tasks present participants with part of an inference (often as part of a truth table) and invite them to suggest conditions that would complete the inference. For example, Evans (1972a) gave participants a conditional such as ‘If there is a red triangle on the left then there is not a green square on the right’ and sixteen cards representing four different shapes each of four different colours. They were either instructed to construct a case which made the conditional true or instructed to construct a case which falsified the conditional and were expected to take two cards and place them together in an appropriate form. So, to show a case in which the example rule was true, the card representing the red triangle might be placed to the left of the card representing the blue circle. To falsify the rule, the participant could place the card representing the green square to the right of that representing the red triangle. Similar materials are also used in an alternative form of evaluation task which presents participants with a rule and stimulus that relates to the rule. Participants then say whether the stimulus confirmed or falsified the rule (e.g. Marcus & Rips, 1979).

Selection tasks provide a participant with options that they should choose to meet the terms of an inference. For example, Wason’s (1968) selection task provided participants with four cards and told them that each card had a letter on one side and a number on the other. Two of the cards had letters on the sides facing the participant (e.g. ‘E’ and ‘K’) and the other two had numbers (e.g. ‘3’ and ‘8’). Participants were given a conditional relating what was on each side of a card (like, ‘if there is an E on one side

then there is a 3 on the other’) and asked to select the cards that they would need to turn over to prove that the conditional was true or false. Using classical logic, you need to check the cards which could potentially falsify the conditional to show that it is true (Popper, 1959). In this example, classical logic suggests that participants should select the ‘E’ card and ‘8’ card. Each of these cards would falsify the conditional if they had a ‘3’ or ‘E’ respectively on their reverse sides.

Logical terms are normally used to define conditional sentences. Researchers might, therefore, expect participants to provide a logical response to each of these tasks. However, studies have often shown systematic biases in responses which do not reflect expectations based on classical logic.

## **1.2. Systematic Biases in Conditional Reasoning**

If human reasoning in conditional inference evaluation tasks followed the tenants of classical logic, we would expect to see people consistently endorse MP and MT inferences and consistently reject DA and AC inferences. However, this pattern is rarely observed in experimental results. In a meta-analysis of experimental evaluation tasks, Schroyens, Shaeken and d’Ydewalle (2001) found that, while endorsement of abstract MP inferences did approach the ceiling (96.8%, as expected based on classical logic), MT inferences were endorsed only 74.2% of the time. Schroyens et al. (2001) also showed that DA inferences were endorsed more than half of the time (56.0%) and endorsement of AC inferences (64.0%) approached the level of MT endorsement – despite classical logic suggesting that DA and AC endorsements should approach zero. Therefore classical logic cannot fully account for all the biases that people demonstrate biases in responses to reasoning questions.

Similarly, in Wason’s (1968) selection task, only a minority of participants routinely choose both cards required to test the truth of the rule and no others. Oaksford and Chater (1994) conducted a meta-analysis of studies using the selection task and found that while 89% chose the card equivalent to the ‘E’ card in the above example, only 25% chose the also required card equivalent to ‘8’ in the example. They also found that 16% of participants would turn over the card equivalent to ‘K’ and 62% would turn over the card equivalent to ‘3’. This result could be an example of confirmation bias – where



people select cards that might provide confirmatory evidence of the conditional rather than the cards which could falsify it (as is required to show the rule is true).

### **1.2.1. Biases Involving Negation**

There are also two well-established effects that are related to the negation of propositions in conditional reasoning tasks: matching bias (Evans, 1972a; Evans & Lynch, 1973); and, negative conclusion bias (Evans, 1977).

#### **1.2.1.1. Matching Bias**

Evans (1972a) first noticed that people showed a bias in participants' response to conditional reasoning questions towards answers that matched the terms used in the question. Evans (1972a) conducted a truth table task in which participants were provided with a condition rule and were asked either to generate an example showing the rule is true or an example falsifying the rule. The task used rules like 'If there is a red triangle on the left, then there is not a green square on the right'. Participants saw a 4x4 matrix of figures of different shapes (circle, triangle, cross, square) and colours (red, yellow, green, blue) from which to generate responses. Participants showed a tendency to select coloured shapes named in the rule (so, in the example given the red triangle and green square) and use them in their responses. For example, almost all participants provided responses intended to verify all of the rules which reflected the terms of the rules. So, the coloured shape described by the rule where the rule used an affirmative proposition and a different coloured shape where it used a negative proposition (e.g. a red triangle on the left and blue circle on the right being used to verify the example rule above). The only exception to this in the verification tasks were two participants (out of 24) who verified rules of the form 'if not  $p$  then not  $q$ ' with responses representing propositions  $p$  and  $q$  (which is also consistent with the rule being true). This tendency to match the terms of the rule was more pronounced when participants gave examples to falsify the rule. A correct falsification of the rule requires the choice of a coloured shape consistent with the antecedent and another inconsistent with the consequent (e.g. a red triangle on the left and green square on the right would correctly falsify the example rule above). Participants were significantly less likely to provide a correct falsification when the antecedent was negative (i.e. they picked examples that matched the positive version of the antecedent). Participants were significantly more likely to provide a correct example falsifying the rule where the

consequent was negative (i.e. they picked examples that matched the positive versions of the consequent).

Matching bias has also been observed in tests based on Wason's (1968) selection task paradigm. Evans and Lynch (1973) conducted an abstract selection task experiment. As with Evans' (1972a) truth table task, this task used abstract conditional rules (of the form 'if  $p$  then  $q$ ') which systematically presented affirmative and negated antecedent and consequent terms. As with all selection tasks, they provided participants with cards representing the true antecedent ( $p$  where the antecedent is affirmative and  $not-p$  where it is negative), the false antecedent ( $not-p$  and  $p$  where the antecedent is affirmative and negative respectively), the true consequent ( $q$  and  $not-q$  where the consequent is affirmative and negative respectively) and the false consequent ( $not-q$  and  $q$  where the consequent is affirmative and negative respectively). They anticipated a matching bias in which participants would be more likely to choose the cards representing  $p$  and  $q$  in preference to the cards representing  $not-p$  and  $not-q$  respectively. This drove the specific predictions that: the card representing the true antecedent would be picked more where an antecedent was affirmative than where it was negative; the card representing the false antecedent case would be more selected where the antecedent was negative than where it is affirmative; the card representing the true consequent would be picked more where a consequent was affirmative than where it was negative; and, the card representing the false consequent case would be more selected where the consequent was negative than where it is affirmative. Their results were consistent with these predictions. Similar to Evans (1972a) truth table results, Evans and Lynch's (1973) results suggest a bias towards the selection of cards which match affirmative versions of the conditional's terms. This bias shows that participants are not responding in a logically equivalent way to each condition.

Subsequent research has shown that the matching tendency in truth table and selection tasks are robust phenomena (Evans, 1998). Using conditional rules of the form ' $p$  only if  $q$ ' or ' $q$  if  $p$ ' instead of 'if  $p$  then  $q$ ' also yields matching bias (Evans, Clibbens & Rood, 1996). However, no similar matching effect is observed in disjunctive reasoning which uses rules like 'It is either  $p$  or  $q$ ' (Evans & Newstead, 1980). Experiments have shown that naturalistic materials, like 'If I finish my work, then I'll be home in time for dinner', can reduce or remove matching bias (Oaksford & Stenning, 1992) although

matching bias is still observed where the naturalistic materials lack context (e.g. Manktelow & Evans, 1979).

These observations of a matching bias show that people do not respond to conditional reasoning tasks using rules with negation in a logically equivalent way to tasks using entirely affirmative rules. For example, classical logic would require people to falsify the conditionals given in Evans' (1972a) truth table task by giving examples that affirmed that antecedent and contradicted the consequent. Participants were significantly less likely to correctly falsify the conditional where antecedents were negative and more likely to correctly falsify the rule where consequents were affirmative suggesting a tendency to match terms in a manner inconsistent with classical logic.

#### 1.2.1.2. Negative Conclusion Bias

Roberge (1971) used conditional inference evaluation questions which asked whether given inferences (each with an abstract rule, minor premise and conclusion) were valid. The rules systematically varied the negation of the antecedent and consequent terms. The minor premise and conclusion either created a valid MP inference or an incorrect inference form with the opposite conclusion (of the form  $p \rightarrow q, p \therefore \neg q$ ). Roberge's (1971) results showed more errors when participants were expected to deny an inference of the form  $\neg p \rightarrow q, \neg p \therefore \neg q$  than the logical equivalent with an affirmative conclusion ( $p \rightarrow \neg q, p \therefore q$ ). Similarly, participants made more errors where they were expected to endorse an affirmative conclusion in inferences of the form  $\neg p \rightarrow q, \neg p \therefore q$  than the logically equivalent form with a negative conclusion ( $p \rightarrow \neg q, p \therefore \neg q$ ). These results suggest that people are more inclined to endorse a conclusion with a negation in it than the logically equivalent affirmative conclusion. A similar bias was demonstrated by Evans (1972b) using MT and AC inferences in an evaluation task. Evans (1972b) presented conditionals about letters and numbers with affirmative and consequent polarities systematically and minor premises related to the consequent. For example, forms like 'If the letter is not G then the number is 9' and 'Not 9' were used to describe the conditional  $\neg p \rightarrow q$ , and minor premise  $\neg q$ . Participants were given a choice of  $p$ ,  $\neg p$  and *indeterminate* and asked which represented the appropriate conclusion (e.g. in the example, the choices would have been 'G', 'Not G' and

‘indeterminate’). Evans (1972b) found that the conclusion required to endorse MT inferences was selected significantly less where the conclusion required was affirmative than where it was negative. For example, participants selected  $\neg p$  as a response to  $p \rightarrow \neg q$ ,  $q$  and  $p \rightarrow q$ ,  $\neg q$  more frequently than they selected  $p$  as a response to  $\neg p \rightarrow q$ ,  $\neg q$  and  $\neg p \rightarrow \neg q$ ,  $q$ . Similarly, the conclusion required to endorse AC inferences was selected significantly more where the conclusion required was negative than where it was affirmative. So, participants selected  $\neg p$  as a response to  $\neg p \rightarrow \neg q$ ,  $\neg q$  and  $\neg p \rightarrow q$ ,  $q$  more frequently than they selected  $p$  as a response to  $p \rightarrow q$ ,  $q$  and  $p \rightarrow \neg q$ ,  $\neg q$ .

Evans (1977) investigated this apparent bias using all four types of conditional inferences (MP, MT, AC and DA) with abstract conditional rules that systematically varied the polarity of the antecedent and consequent terms (i.e. ‘if  $p$  then  $q$ ’, ‘if not  $p$  then  $q$ ’, ‘if  $p$  then not  $q$ ’ and ‘if not  $p$  then not  $q$ ’). Each trial asked a participant in one of two groups (one group used rules in the ‘if  $p$  then  $q$ ’ form while the other used the ‘ $p$  only if  $q$ ’ form) whether they would endorse the conclusion which followed the rule and a minor premise. There were no differences between the polarity conditions in the level of MP inference endorsement, which appeared at ceiling. AC, MT and DA inferences were more likely to be endorsed when the conclusion was negative for the group that used the ‘if  $p$  then  $q$ ’ form. However, there were no differences in AC, MT and DA inference levels where participants saw the ‘ $p$  only if  $q$ ’ form.

Further research has shown that this observation of negative conclusion bias is robust (e.g. Pollard and Evans, 1980). Evans, Clibbens & Rood (1995) used large samples of participants (at least 42 in each group) to demonstrate a highly significant bias for DA and MT inferences (although not MP or AC where endorsement responses often approached the ceiling). They observed this effect when using all of the ‘if  $p$  then  $q$ ’, ‘ $p$  only if  $q$ ’ and ‘ $q$  if  $p$ ’ forms of the conditional. Evans, Clibbens & Rood (1995) found no evidence that this negative conclusion bias reduced when they used naturalistic materials.

The consistent observation of a negative conclusion bias in conditional inference evaluation tasks suggests that the use of negation does have an impact on participants’

responses. Whether participants find a negative conclusion easier to understand or harder to deny (Evans, 1972b) than the equivalent affirmatives, it is clear that people have a complex interpretation of negative form which go beyond the single, simple operator suggested by classical logic.

### **1.2.2. Conditional or Biconditional?**

Conditional statements could be confused with biconditional statements. A biconditional rule (written as ‘if and only if  $p$  then  $q$ ’ or in logical form ‘ $p \leftrightarrow q$ ’) functions as the intersection of two conditionals: (i) ‘if  $p$  then  $q$ ’ and (ii) ‘if  $q$  then  $p$ ’. If a biconditional relationship is used in place of a conditional relationship between two propositions: an AC inference of the form (i) is equivalent to a MP inference of form (ii); and, similarly, DA with (i) is equivalent to MT with (ii). Therefore, both AC and DA inferences are valid under a biconditional interpretation of a conditional. It is, therefore, possible that the endorsement of AC and DA observed in previous studies could be because some participants interpret the conditional statements as biconditional statements.

Taplin and Staudenmayer (1973) observed that 64% of their participants made consistent conditional inference evaluation judgements in line with a biconditional interpretation of a conditional rule. However, when participants were allowed to indicate that a conclusion was ‘always true’, ‘sometimes but not always true’ or ‘never true’ (rather than having to label a conclusion as ‘true’ or ‘false’), only 15% consistently gave responses implying a biconditional interpretation. This second task increased the number of participants consistently responding in line with propositional logic from 3% in the earlier task to 36%. However, the majority of participants were still not consistently responding according to the tenants of propositional logic which suggests that systematic biases are not simply due to a substantial proportion of participants making biconditional interpretations.

Subsequent studies have confirmed that, while biconditional interpretations of conditional relationships can be made (e.g. Markovits, 1988), conditional reasoning biases are not simply caused by a proportion of participants making biconditional interpretations of conditional rules (e.g. Wildman & Fletcher, 1977; Marcus & Rips, 1979; Romain, Connell & Braine, 1983). A more sophisticated account of human

reasoning is therefore required to explain the discrepancy between participants' responses to reasoning tasks and the responses predicted by classical propositional logic.

### **1.3. Accounts of Reasoning**

Researchers have proposed a range of models to explain reasoning biases (Manktelow, 2012). These include accounts which regard classical logic as the normative standard for human reasoning. These explain biases in terms of characteristics of the algorithms that the accounts suggest are used in human reasoning to approximate classical logic. Such accounts include those based on mental logic processes (e.g. Braine & O'Brian, 1991, and Rips, 1994) and the representation of propositions using mental models (Johnson-Laird, Byrne & Shaeken, 1992). Other accounts suggest that a different normative standard should be used to understand human reasoning (such as probability theory, Oaksford & Chater, 2007). Finally, some models combine different systems or processes in an account of human reasoning (such as Evans', 1984, 2006, heuristic-analytic or dual-process approach).

Any account of people's reasoning will seek to account for observed results. The model can be evaluated based on its power to explain the observations considered in its development and its ability to predict further results. Beyond this, a full account of reasoning should be coherent and plausible. Marr (1982) set out three levels on which we can understand complex information-processing systems: Computational; Algorithmic; and Implementation. A computational level account provides the theoretical goal of a process and explains why it is appropriate. An algorithmic (or representational) account considers how the computational theory is implemented through a representation of the input and output and algorithm through which the transformation is achieved. Finally, an implementation account shows how the algorithmic account is physically realised. Marr (1982) argued that a complete account of a complex information-processing system in cognitive psychology must address all three levels.

Algorithmic-level accounts of cognitive processes are required to show how people may understand inputs and then process them to deliver an output. Most accounts of reasoning seek in the first instance to provide a model at the algorithmic level.

People physically process information in their brains and therefore a full account of human information processing should ultimately address the brain's operation. However, an account of cognitive processes at the biological level is complex and a lack of a full biological explanation should not preclude the proposal of algorithmic and computational level accounts. In the absence of sufficient information about how the brain implements a process, the proposed model should at least be plausible given what we know about operations within the brain.

A computational level theory may not be as obviously necessary as an implementation and algorithmic account, but it is important to understand what an information-processing activity is seeking to achieve. A computational account provides an account of why a process exists where the algorithmic account explains how the process could work. We can only judge the effectiveness of a process against a standard which reflects the process's purpose. A computational account should motivate an algorithmic account. Without a computational account, we cannot see whether an algorithmic account provides consistency or fully understand the process. For example, a heuristic might provide a good match to some experimental data but, without clarity of what motivates the reasoning processes that underpin the data, we cannot tell whether the heuristic provides an accurate algorithmic account. It may be that the heuristic matches some data by chance but may extend to other tasks. Oaksford and Chater (1995) have sought to apply Marr's (1982) framework to reasoning, arguing that deference to classical logic is inadequate as a computational-level account.

### **1.3.1. Mental Logics**

Some accounts of reasoning use classical logic as both the computational level account and also the basis of their algorithmic model (e.g. Braine & O'Brien, 1991, and Rips, 1994). These approaches suggest the human cognitive system implements a classical logic at the algorithmic level, but memory and processing capacity constraints limit this system. The resulting divergences between internal logic and classical logic explain reasoning biases like those discussed above (O'Brien & Manfrinati, 2010).

Braine and O'Brien (1991) provide a model for the treatment of conditional statements based on a mental logic. The model uses inference schema supposedly linked to the

lexical entry for '*if...*'. This schema provides a rule for MP and the schema for conditional proof. The schema for conditional proof states that the form '*If  $p$  then ...*' is derived and evaluated by assuming  $p$  and allowing the assertion of '*if  $p$  then  $q$* ' for any proposition  $q$  that is suggested by other assumed information. In this mental logic, Braine and O'Brien (1991) suggest that the rule for MP is applied straightforwardly. However, other inferences require a more sophisticated process which combines the schema for conditional proof with other inferences. For example, Braine and O'Brien (1991) suggest that MT inferences can be verified as correct either through a simple process (mistakenly interpreting the conditional as a biconditional – see §1.2.2 – and applying the MP process) or a complex process (constructing a reductio ad absurdum argument – e.g. to evaluate  $p$  when told that '*if  $p$  then  $q$* ' and ' *$q$  is false*' you would need to reason "if  $p$  were true, then  $q$  would have to be true, but it is not, so  $p$  must be false"). However, the usual intermediate process in their model which would be applied where people reject a biconditional interpretation but do not develop the complex argument is inconclusive (which would restrict MT endorsement). The need for a complex process to correctly solve MT problems (unless biconditionality is incorrectly assumed), explains why endorsement levels are typically much lower than for MP – even though both are logically correct.

Similarly, Rips (1994) proposes the 'Psychology of Proof' (PSYCOP) model of reasoning which provides for a series of rules and a system to apply them. For example, a rule called 'Backward IF Elimination' is equivalent to MP. Inconsistencies in responses to reasoning tasks occur because some rules are more likely to be applied than others. As with Braine and O'Brien's (1991) approach, there is no single rule equivalent to MT.

Both of these models, therefore, have an explanation for the basic bias observed in conditional inference evaluation task, the low endorsement of MT compared to MP despite both being equally logically valid. Similar biases can also be explained using these models through limitations and constraints on the internal logic that does not apply to classical logic (Braine & O'Brien, 1991; Rips, 1994, chapter 5). However, by assuming that conditional reasoning tasks are treated as problems to be solved using internal logical processes, these models have a harder time accounting for differences in responses to logically equivalent tasks. For example, how would the same flawed



internal logic create context effects like the tendency for responses to naturalistic materials to be more in line with classical logic than for responses to abstract materials (Johnson-Laird & Byrne, 2002)? It is similarly unclear how these models could account for the systematic biases observed when using negations.

Noveck and O'Brien (1996) have sought to address these questions by proposing that, in some contexts, pragmatic reasoning schemas interact with Braine and O'Brien's (1991) model. Cheng and Holyoak (1985) proposed that people reason using 'pragmatic reasoning schemas', abstract knowledge structures derived from experiences such as 'permissions', 'obligations' and 'causations'. For example, a permission schema would apply where meeting a precondition was required before an action could be carried out. Cheng and Holyoak (1985) suggested that a permission schema, for example, would suggest the following rules relating antecedent and consequent terms in conditions as a precondition and action respectively:

- If the action is taken, then the precondition must be satisfied.
- If the action is not taken, then the pre-condition need not be satisfied.
- If the precondition is satisfied, then the action may be taken.
- If the precondition is not satisfied, then the action must not be taken.

Cheng and Holyoak (1985) conducted two experiments with conditions that emphasised the potential permissive meaning of conditional rules or provided the conditional rules without that emphasis. Participants then used the conditional rules in tasks using Wason's (1968) selection paradigm. The emphasis on the permissive meaning significantly increased the number of participants selecting the logically correct cards (61% selected the logically correct cards in their second experiment when the permissive meaning was emphasised in a task with abstract materials compared to 19% when the task used more traditional framing). This is in line with participants making use of a permission schema as hypothesised. Noveck and O'Brien (1996) criticised the task and suggested that the enriched features of the pragmatic schema task and relatively impoverished control confounded the results. Their experiments suggested that the permission schema itself did not account for much of the improved performance demonstrated on the task. However, they did suggest that in some contexts such schema may provide participants with the ability to evaluate more inferences correctly according to classical logic than Braine and O'Brien's (1991) mental logic process

alone. Rips (1994) also attempts to address negation effects within the PSYCOP model. He suggests additional rules for the processing of negation that interact with the other rules to produce biases. The use of pragmatic schemas or additional rules potentially extends mental logic accounts to cover content, context and negation effects. However, it remains the case that these theories' greatest predictive power in conditional reasoning tasks is over the more basic biases (like higher endorsement of MP than MT).

The internal logical processes within mental logic accounts provide an algorithmic-level explanation of reasoning processes. At the computational-level, they all assume that reasoning aims to emulate classical deductive logic. One could, therefore, describe reasoning biases as flaws within the representation and processing of the task which constrain people's ability to carry out full logical reasoning.

### **1.3.2. Mental Models**

The mental models approach (Johnson-Laird, Byrne & Schaeken, 1992) recognises that classical logic may not be the most appropriate model for everyday reasoning. As its computational-level aim, mental models theory assumes that people reason by using the meaning of premises, and general knowledge, to imagine the possibilities under consideration (Johnson-Laird & Byrne, 2002). At the algorithmic level, it sets out how these possibilities are represented by the 'models' people generate.

The mental models over which this approach suggests people reason each represents a possibility or set of possibilities in which each of the assertions in the model is true (Johnson-Laird & Byrne, 2002). Each model represents a clause in the premises only when the clause is true within the possibility set out in the model. A model is analogous to a diagram with structures which represent the structure of a particular possibility (although models can also be used to represent possibilities which consist of abstract concepts that cannot be visualised, Johnson-Laird, 2008). Creating models consumes time and retaining them in working memory consumes limited capacity. Therefore, the mental models theory anticipates that inferences requiring fewer models will be more readily made and endorsed than inferences requiring the evaluation of models representing all possibilities.

The mental models approach anticipates that a person doing a conditional reasoning problem will initially consider models which satisfy the antecedent terms of a rule (Johnson Laird & Byrne, 2002). Given a rule of the form ‘if  $p$  then  $q$ ’, a person will first generate a model which represents the possibility of ‘ $p$  and  $q$ ’. This model is represented as follows (the squared brackets indicate that  $p$  is ‘exhausted’ and that it cannot appear in any other models because it cannot appear without  $q$ ; the ellipsis indicates that other models, not yet generated, are possible):

[p]	q
...	

This initial model alone would allow a person to make or endorse a MP inference because it shows the only available possibility when  $p$  is true is one where  $q$  is also true. Someone using this model at a very simple level – for example, to avoid the time and memory resources required to generate further models – may not take into account that  $q$  has not been exhausted and make or endorse an AC inference. This initial model does not provide enough information about the available possibilities to facilitate MT inferences. Someone wanting to test an MT inference would need to generate the ‘implicit models’ which represent the other allowed possibilities. The initial and implicit models are as follows:

[p]	q
$\neg p$	$[\neg q]$
$\neg p$	q

Someone using the initial model and the first implicit model in a primitive way (not reflecting that  $\neg p$  is not exhausted by the first implicit model) could also make DA inferences. The final model needs to be generated to show that DA and AC are false.

Mental models theory, therefore, provides a plausible account of lower MT than MP inference as MP inferences can be drawn using an initial model while MT inferences require an implicit model. The theory also shows how a person who did not fully spell out all possibilities in models could make AC and DA inferences.

Mental models theory integrates ‘the principle of semantic modulation’ and the ‘principle of pragmatic modulation’ (Johnson-Laird and Byrne, 2002). The former principle allows the meaning associated with antecedent and consequent clauses, and information known about the link between them, to affect the development of models. The latter principle states that contextual knowledge related to propositions in a conditional is normally represented in explicit models. These principles both propose that semantic and pragmatic information can be added to models, suppress the generation of otherwise acceptable models or enable the construction of a full set of models. For example, given the conditional ‘if you turn the key, then the car will start’ and the contextual information, ‘if there is no fuel the car cannot start’, the generation of a model representing the possibility ‘you turn the key, there is no fuel and the car starts’ will be suppressed. Mental models theory uses these principles to account for a range of content and context effects.

Johnson-Laird and Byrne (2002) also seek to explain the commonly observed negation effects using mental models theory. They suggest that negative conclusions are easier to derive from mental models than affirmative conclusions (which requires resolution of double negations occurring in the models) which may account for negative conclusion bias. They seek to explain the observed matching bias on the basis that it is hard to understand one that assertion (like “The number is nine”) refutes another (like “The number is four”). As mental models represent true assertions, people tend to select items based on the terms in the initial mental model – and not necessarily items which reflect assertions that should refute those terms.

The mental models approach can, therefore, provide a plausible account for a diverse range of experimental data (Johnson-Laird, 2008). However, there remain questions that it cannot address. For example, given the apparent constraints on quick model generation and model retention, how do people apply the principles of semantic and pragmatic modulation which appear to require consideration of many more models – particularly if subtle differences in interpretation need to be addressed?

The mental models approach also provides an algorithmic account of reasoning that assumes that at the computational level people reason over possibilities generated using

information in front of them and wider knowledge. The approach explains how capacity constraints lead to responses that reflect different experimental findings. However, it is not clear how the computational level aim would motivate this quite specific algorithmic approach. The theory draws on some detailed principles and other caveats on different occasions to explain different results. There would also appear to be a risk with this approach that the theory is insufficiently constrained in its flexibility because a computational level theory cannot motivate the algorithmic model. It could then be adapted to address different results in ways that may have limited computational justification.

### **1.3.3. Probabilistic Approach**

Classical logic typically deals with propositions and inferences that are certain and indefeasible. However, people often deal with uncertain situations and it may often be right to overturn a prior conclusion given new information. Oaksford and Chater (1991) argued that the need to allow for uncertainty and defeasibility suggested that that classical logic may not provide the best computational level theory of human reasoning. The probabilistic approach, therefore, proposes probability theory – which provides a calculus for uncertainty rather than indefeasible conclusions based on full information – as a more suitable model for a computational level theory of human reasoning (Oaksford & Chater, 2001).

Under the probabilistic approach, responses to reasoning problems that are apparently illogical occur because participants use reasoning strategies which reflect probability theory (which better suits their everyday experiences) rather than classical logic. For example, the probabilistic approach suggests, in responding to the selection task, the inference process may aim to pursue the best chance of getting the greatest reduction of uncertainty about the truth of the rule (Oaksford & Chater, 1994). This means that, rather than using an approach to reasoning that aims to yield the correct answer in classical logic, people use an approach that is likely to provide the greatest information. Oaksford and Chater (1994) supported this theory by showing that a model based on optimal data selection predicted matching bias and provided a good fit for data from previous selection task studies.

Oaksford, Chater and Larkin (2000) extended the probabilistic account to cover conditional inference tasks. They predicted that the level of endorsement of a given inference would correlate with participants' perceived likelihood that the inference is true, given what they knew. Therefore, their model considers how people would assign probabilities to propositions  $p$  and  $q$  used in a rule of the form 'if  $p$  then  $q$ '. The key parameters required in the model are the conditional probability of  $q$  occurring given  $p$  has occurred (written ' $P(q|p)$ '), the probability of  $q$  not occurring (written ' $P(\neg q)$ ') and the probability of  $p$  occurring (written ' $P(p)$ '). The model sets the following parameters:

$P(q|p) = a$       This can be represented as ' $P(q|p) = 1 - \epsilon$ ' where ' $\epsilon$ ' represents the probability of exceptional circumstances under the conditional where  $q$  does not occur despite  $p$  occurring.

$P(\neg q) = b$

$P(p) = c$

Using these parameters, the model includes conditional probabilities derived for each inference.

MP       $P(q|p) = a$

DA       $P(\neg q|\neg p) = (b - (1 - a)c) / (1 - c)$

AC       $P(p|q) = (ac) / (1 - b)$

MT       $P(\neg p|\neg q) = (b - (1 - a)c) / b$

Oaksford et al. (2000) conducted several experiments in which the probabilities of propositions in conditional inference tasks were varied. Responses were consistent with the probabilistic model's predictions.

As with the optimal data selection account of the selection task, the probabilistic account of conditional inference predicts the commonly observed negation effects. These predictions use the assumption that a negated proposition has a much higher probability of being true than the non-negated original proposition. The contrast class account of negation, which will be discussed further below (Oaksford & Stenning, 1992), informs this assumption.

The probabilistic approach, therefore, provides a computational level account that would predict the standard reasoning biases and effects associated with negation in conditional reasoning tasks. It is also easier to account for revised understandings of the likelihood of propositions using probability logic than classical logic so content and context effects can be explained as functions of the computational level theory rather than necessitating the introduction of additional representational principles, schemas or heuristics (Oaksford & Chater, 2007).

Accounts that treat classical logic as the computational level model have well developed algorithmic theories which the probabilistic account initially lacked. However, the development of Causal Bayesian Networks applied to artificial intelligence has provided an algorithmic account which applies probabilities to reasoning (e.g. Pearl, 2000; Sloman, 2005). An algorithmic account that uses Causal Bayesian Networks and is consistent with the probabilistic account may better predict reasoning responses than the mental models approach (e.g. Ali, Chater & Oaksford, 2011).

#### **1.3.4. Multiple Processes**

Some accounts of human reasoning and the observed reasoning biases suggest that it may involve the interaction of two or more processes (Stanovich & West, 2000).

One dual-process approach is the Heuristic-Analytic Theory proposed by Evans (1984). This theory suggests that reasoning uses an analytical process capable of drawing inferences but that this is dependent on initial pre-attentive heuristic processes. These heuristic processes select ‘relevant’ information. Matching bias is therefore explained using this theory as caused by the heuristic processes identifying matching cards as relevant rather than as a failure of analytic reasoning (Evans, 1984). Evans (2006) updated this theory, replacing the simple serial structure of the earlier version with greater interaction between analytic and heuristic systems. The analytic processes are still dependent on heuristic processes but three principles are applied which mediate those processes: the singularity principle – that people only consider one hypothetical possibility at a time; the relevance principle – people consider the most relevant model of the current context; and the satisficing principle – that models are accepted if they satisfy current goals. Evans’ (2006) revised theory still accounts for matching bias with

reference to heuristic processes which satisfy a person's need for an answer before completing a full analysis. However, it better reflects content and context effects through its additional principles.

Other dual-process models make a similar separation of processing systems. Sloman (1996) divides reasoning processes between an associative system (which provides outputs based on similarity and temporal structure) and a formal system based on logical rules. Epstein (1994) provides a model which incorporates psychoanalytic ideas by proposing separate experiential and rational systems.

Stanovich and West (2000) reviewed a range of dual-process approaches. They suggested that many models made a distinction between an associative, automatic and quick 'System 1' which places little burden on cognitive resources and an analytic, controlled and slow 'System 2' which requires a high level of cognitive resources. However, Stanovich and West (2000) found less commonality in the mechanisms different models proposed for the interaction between systems. Some suggest that the two systems operate in serial with one process dependent on the other (e.g. Evans, 2006) while others suggest that both systems operate in parallel – simultaneously generating different responses to reasoning problems (e.g. Sloman, 1996). Subsequent research has considered the relationship between System 1 and System 2. Thompson's (2010) metacognitive dual-process theory suggests that System 1 may provide an initial response to a given task before System 2 is activated and either supports or revises the response. This interaction is dependent on a metacognitive feeling of rightness – the extent to which an answer is sufficient or further analysis is required – and judgement of confidence – an evaluation of confidence in the answers derived – to determine how System 2 is engaged. Verchuren and Schaeken (2010) look at how a dual-process framework can integrate several different accounts of reasoning. Their model sets out three levels at which dual-processes (a System 1 and a System 2) may be distinguished: decisions to incorporate background information; where background information is needed; two ways of taking background information into account (they suggest that probabilistic and mental models approaches suggest heuristic and analytic systems at this level); and, where counterexample information is needed, two approaches to counterexample retrieval.



Multiple process models of reasoning normally provide algorithmic-level accounts of reasoning which suggest complex computational-level accounts that are trying to meet multiple goals. For example, Evans' (1984, 2006) model suggests that people are aiming to respond logically to reasoning tasks but that other factors, such as quick response time, are also important. Verchuren and Shaeken (2010) appear to suggest that different types of logic may be useful at different times (particularly if you treat their dichotomy between probabilistic and mental models approaches as a computational-level distinction between probabilistic and propositional logic aims). Superficially such an integrative approach has some attraction: People can reason using formal propositional logic when required to but this is a slow and deliberate process that experiments have suggested is not engaged in everyday reasoning. Even scientists appear to routinely use other processes which demonstrate systematic biases from the correct responses under propositional logic (e.g. Kern, Mirels & Hinshaw, 1983). However, it may also simply be that deliberative reasoning with formal logic is a special case of higher cognitive functions which uses processes unconnected with everyday reasoning and is therefore not helpfully integrated into a single model.

### **1.3.5. *Alternative Models***

Some other models have also been proposed to account for systematic reasoning biases. These include Cheng and Holyoak's (1985) account based on the use of pragmatic reasoning schemas and Politzer and Bonnefon's (2010) proposal of a calculus of possibilities.

Cheng and Holyoak's (1985) approach has already been discussed above as Noveck and O'Brien (1996) sought to partially integrate it into Braine and O'Brien's (1991) mental logic model to help account for context effects. Cheng and Holyoak (1985) suggest that pragmatic schemas are domain specific inference forms (a set of rules which govern inferences which are selected based on context). When activated, these schemas have the effect of applying a different 'logic' depending on whether the participants aim to respond to a conditional as representing a permission, causation or obligation. This suggests an alternative algorithmic-level view which recognises that people may approach different problems with different computational-level aims. However, it ultimately seeks to distil the processing of reasoning problems into the implementation of logical rules like other approaches based on internal mental logics.

Politzer and Bonnefon (2010) suggested that neither the logics of propositions nor probabilities are adequate to explain the basis for human reasoning. They suggested that a calculus of possibilities may better account for all observed phenomena. Such an approach would place substantial weight on the level of plausibility a person assigns to a task's premises and conclusions (under this proposal plausibility is an indication of the credibility attached to a conjecture on the basis of its source's reliability). This is an interesting idea which requires further testing.

#### **1.4. Accounts of Negation**

As shown above, the concept of negation has been widely used in conditional reasoning experiments and associated with the often observed matching and negative conclusion biases. However, many experiments have simply used the word 'not' to negate a statement and assumed that it functions as the negation function in propositional logic (a denial). Some other experiments have used a form of implicit negation (which asserts an affirmative proposition which contradicts a presupposition – an otherwise assumed or suggested proposition – thereby negating it). These two forms of negation may have different implications, but the reasoning literature rarely considers these. The study of reasoning has not considered other forms of negation which may frequently occur in natural language.

Propositions involving negation may also be understood differently depending on the context, form of the negation and content. People use different negative operators in different ways to specify the scope of the negation and create different implications. Sometimes the scope of the negation may be ambiguous and sometimes it may be specific. Cognitive processes may also consider propositions with negation differently from affirmative propositions (for example, some propositions with negation take longer to process and understand, Clark, 1974). Therefore, consideration of the role of negation in logic and language is needed to understand the role of negation in reasoning processes.

##### **1.4.1. Logical Operator**

Modern formal logic is based on the operation of propositions (Sommers, 1970). Such logic interprets negation as an operator on a proposition that is external to the

proposition and has the function of denying it (Horn, 1989). So the negation of the proposition ‘The bike is blue’ is treated as ‘It is not the case that the bike is blue’ (a proposition that would allow all possibilities apart from that a given bike is blue – the logical complement of the proposition that the bike is blue).

Logically a given proposition and its negation are mutually exclusive and any given proposition or its negation must be true. In propositional logic, the law of double negation – that the negation of an already negated proposition (e.g. it is not the case that the bike is not blue) is equivalent to the proposition without negation (e.g. the bike is blue) – also applies.

Within the calculus of propositions, treating negation as a logical complement operator is clear and may be sufficient. However, it is hard to apply propositional logic to many of the circumstances that we may face in everyday life. For example, when the truth or falsity of a proposition is irrelevant (such as statements about nonexistent subjects ‘the king of France is not bald’ and ‘the king of France is bald’ or subjects to whom the proposition is not appropriate ‘the number twelve is bald’ and ‘the number twelve is not bald’), unknown (‘Schrödinger’s cat is alive’ and ‘Schrödinger’s cat is not alive’ could both be true or false prior to observation of the cat), indeterminate (such as a reference to the future ‘there will be rain tomorrow’ and ‘there will not be rain tomorrow’) or when seeking to use negation to indicate a more precise oppositional state (such as the contrary operators and approaches to narrow the scope of negation discussed below).

#### 1.4.1.1. Other Logics

Aristotle (trans. 1923) used the logic of terms, rather than the logic of propositions, which recognised that different forms of denial are possible (Englebrechtsen, 1976). Aristotle’s logic of opposition (which provided an early and influential account of negation as a concept, Horn, 1989) used two laws to define its approach to negation: the law of contradiction (*Met.* 1005b19-23) and the law of the excluded middle (*Met.* 1011b23-24). The law of contradiction states that a something cannot be both a thing and not that thing at the same time in the same respect. The law of the excluded middle states that anything must either be or not be. Aristotle defined two types of negation based on these laws: contradictory and contrary negation. The law of contradiction applies to both types of negation (*Met.* 1011b17-19), but the law of the excluded middle

only applies to contradictory negation (*Met.* 1057a34-35). So, contrary negation implies an opposite. A situation could occur that cannot be represented by either an affirmative proposition or its contrary negation (for example the contrary negation of ‘the cat is white’ would lead you to believe that ‘the cat is black’ and so also rule out the possibilities that the cat is grey or ginger). Contradictory negation of a proposition describes all possible states not accounted for by the affirmative proposition (so the negated proposition ‘the cat is not white’ would allow for the possibilities that the cat is black, grey, ginger or any other colour but white).

Since Aristotle, philosophers have proposed a range of multivalued logical systems that seek to create a dichotomy between two types of negation. Horn (1989) lists 27 approaches which distinguish two types of negation. These include: Russell’s (1905) ‘secondary occurrence of description’ (wide scope – the negation applies to a whole proposition) versus ‘primary occurrence of description’ (narrow scope – the negation applies to an element within a proposition); Von Wright’s (1959) weak versus strong; and, several definitions of external versus internal (e.g. Clark, 1974). Scope is discussed further in §1.4.2.1. These approaches all have different definitions of their two types of negation, but Horn (1989) notes some commonalities. For example, one of each pair of negation types (e.g. wide scope, weak or external in the examples above) is generally a propositional operator and more like Aristotle’s contradictory operator. The other operator in each pair (e.g. narrow scope, strong or internal) operates more narrowly, either on or within a predicate (the part of the sentence that modifies the subject) and is more like Aristotle’s contrary operator. Despite these options for alternative understandings of negations, some logicians have continued to maintain that within a two-valued logic (where the only truth values are ‘true’ and ‘false’) only an external reading is necessary (Horn, 1989). Karttunen and Peters (1979) offer a compromise approach that allows a sentence to have a single logical meaning but multiple potential implications which are not truth-conditional. This ambiguity based on the conventional and conversational implications of sentences may be acceptable to some logicians, but it has not been investigated experimentally for people’s everyday reasoning.

If negation is treated as having more than one different implication in different logics, then it must have wider everyday usage than as a complement forming operator creating a proposition symmetrical to its affirmative counterpart. This could affect how people

process negations in reasoning tasks – particularly if they are approaching them naïve to their formal logical implications. In developing reasoning tasks that use negation, we should, therefore, consider what different understandings negation has and what impact these might have on participants' reasoning.

#### **1.4.2. *Natural Language***

Symbolic logics can define and articulate the form and role of negation. In natural language, the roles and forms of negation have evolved and allow for many different articulations of negation with subtle distinctions and different implications. Subtle distinctions in sentences that use negations leading to different implications can create ambiguity in the understanding that the use of negation should provide.

In natural language, negation takes a large number of forms. For example: using words like 'not', 'no' or 'never'; using affixes applied to words like 'un-', 'dis-' or '-less'; and using statements which implicitly deny a prior statement ('Was that a bird?' / 'It was a plane'). These different forms of negation can be used to create more contrary or contradictory interpretations.

The different ways of using negation in language can also carry a range of implications which don't always match the logical treatment of negation. For example, the double negation in a suspect's claim that they "didn't not kill the man" is likely intended to emphasise the denial of the proposition "kill the man". In propositional logic, the double negation would cancel out and so the former statement would be interpreted as an affirmation of the latter proposition. Such use of additional negators to emphasise rather than disregard negation can be seen in many languages and can emerge in languages as they change over time (Mazzon, 2004). Applications of negation can vary between dialects (Mazzon, 2004) and between spoken and written modes (Tottie, 1982, found twice as many instances of negation in spoken English when comparing spoken and written corpora). It seems reasonable to assume that naïve experimental participants interpret logical statements involving negation as they would interpret negation in natural language. Therefore, to understand the effect that negation has on logical reasoning we should consider the different implications that the use of negation in language can create and the potential for ambiguity in interpretation.

#### 1.4.2.1. Scope of Negation

In language, a negative operator can have scope over different elements of a statement. The scope of the negation can be affected by the form, context or intonation of the statement (Horn, 1989). For example, by changing the speaker's emphasis, the negation in the sentence "David did not travel to Manchester by car" can refer to either David not going to Manchester or not going by car (Oaksford & Stenning, 1992). As in this example, with no emphasis added the target of negation could be ambiguous. Experiments have shown that article cues (such as preceding the negated item with 'a' rather than 'the') are more influential than position cues (like the order of nouns) in helping people identify the target of negation in a written sentence (Rumain, 1988).

Klima (1964) explored the ambiguity that can occur when a sentence like 'The bike is blue' is negated as 'The bike is not blue'. The sentence could be intended as a denial of the whole proposition (sentential negation, 'It is not the case that the bike is blue') or as a negation of term 'blue' (constituent negation, 'The bike is not-blue'). Klima (1964) highlighted the importance that context plays in interpreting the negation of such a sentence where there are insufficient cues within the sentence itself. The potential to identify different potential scopes in such negative sentences does affect the logical interpretation of the sentence. For example, Russell (1905) argued that an externally negated interpretation (in this example, 'It is not the case that the bike is blue') would be logically true if there was no bike but the internally negated interpretation (i.e. 'The bike is not-blue') would be false. Some others (e.g. Strawson, 1950), suggest that there is a limit to the extent to which such external negations are possible. Further cues can also be added to clarify the scope of negation (e.g. 'It is not the case that the bike is a blue one' initially appears externally negated but the use of the indefinite article implies that the colour is the scope of negation and it is not true if there is no bike).

Hilton (1995) reviews the effects that conversational inference can have on psychological experiments. This includes consideration of Grice's (1975) maxims of conversation. Grice (1975) suggested that people exchanging information normally cooperate to gain mutual understanding using certain shared assumptions. These assumptions included maxims of quality (be truthful), quantity (provide the least information necessary for understanding), relation (be relevant) and manner (be clear). Where these assumptions are apparently broken, conversationalists are likely to

interpret that as having meaning (e.g. where more information than apparently necessary is provided or where a statement does not seem relevant or unambiguous). Hilton (1995) reviews experimental tasks which may have led participants to draw a different understanding to that the researchers intended. This could make the results misleading. For example, the maxims of quantity and relation may lead experimental participants to assume that all the information that they are given is relevant and should be used in a task. This may lead to a weight being given to irrelevant details in decision making tasks.

Negative sentences, therefore, need careful consideration when developing experiments to make sure the scope of the negation is understood consistently in the same way by all participants.

#### 1.4.2.2. Pragmatic Ambiguity

Horn (1989) suggests an alternative way to interpret some ambiguity with negations. He argued that rather than being logically or semantically ambiguous, descriptive negation is pragmatically ambiguous (a sentence like 'John's killer is mad' has a single semantic interpretation but in different contexts could be understood as 'Anyone who would kill John is mad' or 'The individual that killed John is mad'). Horn (1985, 1989) also describes a use of negation that is metalinguistic rather than descriptive (simply denying a proposition). Such metalinguistic negation can be used to register objection to an utterance on any grounds whatsoever, such as a statement's implication, morphology or even intonation. For example, the statement 'it isn't cool, it's cold' does not deny the assertion 'it is cool' but rejects it because 'cool' is insufficiently descriptive of the weather. When negation is used, this potential for pragmatic distinctions and metalinguistic interpretations further suggests that the interpretation of negative sentences is more complex and involves different processes to the interpretation of affirmative statements.

The possibility that different types of negation – or the same type of negation in different contexts – can create different understandings may affect how people process negation in reasoning problems. Experimenting with different types of negation and observing any differences in responses may, therefore, help us understand how people process reasoning tasks.

### **1.4.3. Contrast Classes**

Contrast classes provide a psychological account of negation based on the philosophical account of negation as ‘otherness’ (Oaksford, 2002). The otherness account suggests that the use of negation implies that something ‘other’ than the item negated is the case. The contrast class account exploits this notion by suggesting that negation activates possibilities that could be true if the negated item is not.

People are likely to have formed cognitive categories into which they can place items they encounter or are aware of into (e.g. Rosch, 1973). For example, most people will recognise that items like ‘red’, ‘blue’ and ‘purple’ belong to a category known as ‘colours’. The set of items someone will have in a given category provides a reference class of items that people can access when discussing a topic. So, when the context of a discussion requires consideration of a colour, a person might come up the example of ‘blue’ rather than ‘seagull’ (which is more likely to belong to sets categorised as ‘birds’ or ‘coastal animals’). Items within sets may be more or less typical of the set (more typical items will have more characteristics in common with other items in the set and fewer in common with items in other sets) and items may belong to more than one categorical set (Rosch, 1973). People can use these reference sets to derive contrast sets for an item. The contrast set of a target item will consist of other items which share characteristics with the target item (and so are included in the same categories as the target item) but which are not the target item (Rosch & Mervis, 1975).

The contrast class account of negation therefore suggests that, when an item is negated, people consider other items in the contrast set of the negated item. A contrast class consists of items that would be relevant in the same context as the negated item which a person considers as possibilities in processing the negation. This class is much smaller than the logical complement of the negated item as it only includes items that the individual considers relevant.

Oaksford and Stenning (1992) used contrast classes in their probabilistic ‘processing negations’ account of Wason’s (1968) selection task. They suggested that the contrast class of a given item is normally more likely than the given item itself (the contrast class representing more possibilities). Their experimental results supported the resulting probabilistic model. However, those experiments and others since (e.g. Oaksford and



Moussakowski, 2004) did not show direct evidence of spontaneous contrast class construction or activation when participants address reasoning problems. Further research is therefore needed to understand whether contrast classes are relevant to everyday reasoning and what circumstances may mediate their use. Asking whether people sometimes activate contrast classes to understand negation in reasoning also suggests a wider question: How do people process sentences with negation?

## **1.5. The Psychology of Negation**

To understand how people use negation in reasoning, we need to understand how people process negation more generally. There has been a long-running debate among philosophers about whether negation is read in the same way as equivalent affirmatives or if it is somehow marked, more complex and less basic (Horn, 1989). In psycholinguistics, a number of effects can be noted when negation is used which suggest an asymmetry between negative and affirmative statements.

### **1.5.1. Processing Time**

Early studies noted that it takes participants longer to understand statements that used negation than equivalent statements without negation (Clark, 1974). For example, Just and Carpenter (1971) asked participants to verify statements like ‘The dots are [red/black]’ and ‘The dots aren’t [red/black]’ with reference to a picture of dots that were either red or black. They found that people responded significantly faster to the affirmative statements than the statements with negation. Such effects would appear to suggest that negation is inherently more complex and takes longer to comprehend than equivalent affirmative statements.

#### **1.5.1.1. Contexts of Plausible Denial**

Wason (1959, 1961) had also observed longer processing times for statements with a negation (such as “Forty-six is not an odd number”) than statements without negation (e.g. “Sixty-five is an odd number”). He also observed that people verified true statements without negation faster than false statements without negation (“Thirty-nine is an even number”). However, he observed the converse with statements containing negations (so a false statement with negation like “Ninety-two is not an even number” yields a quicker response than a true one like “Ninety-three is not an even number”). Wason (1961) suggested that this core difference between affirmative and negative

statements was that negatives are transformed into affirmative equivalents (e.g. “not odd” becomes “even”) taking additional processing time before participants respond. Wason (1965) built on these results with an evaluation task that used statements and related pictures of coloured circles. He observed faster responses where a negated property in the statement might have been expected to be true (for example, saying a single red circle amongst blue circles is ‘not blue’) than where it would not be expected (saying a red circle is ‘not blue’ when most circles are red). Based on these results, Wason (1965) suggests that negative sentences are more easily understood within ‘contexts of plausible denial’. So, understanding of a statement involving a negated item is easier if the negated item could plausibly have been thought true had it not been negated. This may suggest that statements using negation are processed differently depending on whether the reader recognises the appropriateness of the negation.

### **1.5.2. Probing Negations**

MacDonald and Just (1989) looked at what negation did to the activation of the negated word. They conducted a series of probe tasks giving participants a statement including two nouns (conditions were that neither noun was negated or that one or the other was). They then presented a probe word (one of the nouns in the experimental trials – normally an unrelated word in controls) and asked participants to indicate whether the word had appeared in the sentence (in their first experiment) or say the word aloud (in experiments 2 and 3). There was then a comprehension question. They looked at reaction times to the probe word and found that participants responded fastest to either noun when the sentence had used no negation. They found that where a given noun was negated, responses were significantly slower when that noun was the probe than where the other noun was negated. This suggests that negation represses activation of the negated item. In their third experiment, MacDonald and Just (1989) also used words associated with the nouns that were not included in the original sentence as probes. They found that reaction times to associated words were slower where the relevant noun had been negated in the sentence.

Hasson and Glucksberg (2006) looked in more depth at the semantic activation of associated words after the presentation of a statement in which a relevant item may or may not be negated. They conducted lexical decision tasks that started 150ms, 500ms or 1,000ms after the presentation of the priming statement. They found no difference in

responses at the 150ms and 500ms intervals whether the relevant item was negated or not: Participants responded faster where target words were associated with the relevant item and slower where the target work was associated with the opposite of the relevant item. However, at 1,000ms participants responded faster to words associated with the opposite of the relevant item if it had been negated in the priming statement. This suggests that the negated statements were processed in the same way as affirmative statements for up to a second after they were presented. At that point, the negative statements were priming items with opposite associations which the affirmative statements were not. This may be consistent with the contrast class account discussed earlier (see §1.4.3).

## **1.6. Approach**

The use of negation in conditional reasoning tasks has been studied extensively and associated with frequently observed biases. A large number of studies seek to explain the biases. However, there has been relatively little consideration of the different implications that negation can have. While the contrast class account provides a psychological account of negation which goes beyond its role in propositional logic as a complement operator, there is limited direct evidence of contrast class activation in reasoning tasks. Therefore, further exploration of the use of negation in reasoning may be helpful in understanding how people reason and evaluating the proposed accounts.

The current study will investigate the interaction between the use of negation and people's conditional reasoning biases. By considering how negation affects conditional reasoning, it aims to help understand how people reason. As its starting point, this study will look at how different types of negation affect reasoning. Differential effects of different types of negation may suggest that different processes are used to process different types of negation. This study will then consider the role of contrast classes in reasoning. This thesis is organised as follows.

Chapter 2 uses probability rating tasks (Oaksford, Chater & Grainger, 1999) to look at whether using different types of negation affects people's judgement of the likelihood of a proposition. Previous studies (e.g. Oaksford, Chater & Larkin, 2000) have suggested that changing the likelihood of propositions affects how people reason with them.

Chapter 3 looks directly at whether different types of negation affect conditional reasoning. Different types of negation will be used in conditional inference evaluation tasks (based on an expansion of Evans', 1977, design) to see how responses differ.

Chapter 4 will investigate the contrast class account of negation and how it applies to reasoning tasks (testing a hypothesis proposed by Oaksford & Chater, 2007). Reasoning task experiments will explore whether people can reason when explicitly provided with contrast classes and whether they spontaneously use contrast classes when presented with reasoning problems involving negation.

Chapter 5 will consider what conclusions can be drawn from the current study and potential next steps.

## **2. Different Types of Negation and Probability**

Understanding of the interaction between the use of negation and reasoning task performance requires consideration of the nature of negation. Negation can take many different forms and provide many different implications (Horn, 1989). Studies of the use of negation in reasoning have typically used the word ‘not’ to negate the following proposition (e.g. Evans, 1977) or have used an affirmative proposition to implicitly negate another proposition through contradiction (e.g. Evans & Lynch, 1973). Therefore, only a narrow range of forms that negation can take have been used to investigate the effect of negation on reasoning. In this chapter, I look at how other forms of negation in natural language, notably affixal negation (Zimmer, 1964) may affect performance.

Before exploring how different types of negation affect conditional reasoning performance (which is considered in chapter 3), I will provide an initial study of whether different types of negation are used differently in likelihood estimation tasks related to reasoning. People’s perception of the likelihood of given propositions has been previously related to their responses to reasoning tasks using those propositions (Evans, Handley and Over, 2003).

In this chapter, I describe experiments which look at whether people respond differently in probability rating tasks (Oaksford, Chater & Grainger, 1999) when using different types of negation. First, though, I discuss the relationship between reasoning and likelihood estimates and consider some different types of negation.

### **2.1. Reasoning and Probability**

Human reasoning with probabilities, like logical reasoning, is subject to systematic biases. For example, people often violate the conjunction rule (Tversky and Kahneman, 1983) or demonstrate overconfidence in their judgements (Gigerenzer, Hoffrage and Kleinbölting, 1991).

According to probability theory (which respects standard logic), the conjunction rule demands that the probability of two events both occurring must be less than or equal to the lesser probability of either event occurring individually. However, Tversky and

Kahneman (1983) demonstrated that people's probability estimates often violated the conjunction rule, with people often estimating that the conjunction of both events was more likely than either one of the events. They explained their results in terms of representativeness and availability heuristics. As an example of the availability heuristic, Tversky and Kahneman (1983) described an experiment in which people could list more seven letter words ending in "-ing" than with "n" as the penultimate letter (even though the former set is a subgroup of the latter) which affects people's estimates of the frequency of words in those sets. They explain representativeness as overestimation of the probability of events where they conform to a relevant stereotype. This could be a typical example (albeit one that might be narrow and so apply to few examples of the group described). It could also be a diagnostic example (an unusual characteristic but one which is more common in the group in question than the general population, e.g. having divorced four times may be perceived to be more common amongst Hollywood actresses than the general population and is therefore more diagnostic of this group than voting democrat, which may apply to more Hollywood actresses but at a level more typical of the general population).

An overconfidence effect was observed by Gigerenzer et al. (1991) who asked participants a series of general knowledge questions. They showed that participants' estimates of the number of answers they had correctly answered were consistently higher than the number of correct answers they had provided. This implies an overestimation of the likelihood of providing a correct answer. Gigerenzer et al. (1991) suggest that this overconfidence is related to a 'hard-easy' effect in which people show greater overconfidence for more difficult questions (those answered incorrectly more often).

Biases showing people violate the conjunction rule and have over-confidence in some situations might suggest that people are poor at applying probability theory in reasoning. However, there is evidence in these same studies that people can make effective use of frequency data in reasoning and appear to apply the standard calculus of probability. For example, when they asked about the likelihood of one of the conjoined propositions before asking about the likelihood of the conjunction, Tversky and Kahneman (1983) observed fewer instances of the conjunction fallacy. Gigerenzer et al. (1991) eliminated and inverted the overconfidence and hard-easy effects based on their

probabilistic mental models theory. This theory proposes that people iteratively generate internal models to make inferences. These probabilistic mental models use relevant reference classes if the required information for a ‘local’ mental model (like those applied in Johnson-Laird et al.’s, 1992, mental models approach – see §1.3.2) is unavailable. The relevant reference class is the set of all the objects or events of interest in the model. People represent members of the relevant reference class in probabilistic mental models as potential objects connected to cues (wider knowledge about the object). Probabilistic mental models are considered iteratively to identify the objects which have cues that most likely match the requirements of the task. Gigerenzer et al. (1991) attribute overconfidence or underconfidence in a person’s response to the nature of the probabilistic mental models on which it is based (for example, a person will be overconfident if they use probabilistic mental models based on a reference class that seems appropriate to the subject of the question but for which the associated cues do not provide a good match to the detail of the question). Other studies have attributed the effects to different causes. Griffen and Tversky (1992) explored this effect and suggested that overconfidence is mostly observed when the evidence for the answer had high strength (i.e. the forcefulness with which the evidence suggests the answer) and low weight (credibility). They further suggested that an under-confidence effect occurs when the evidence for an answer has low strength and high weight. This suggests that people’s use of probability evidence might be influenced by the context in which is presented (e.g., presentation of information about a small sample – even a single instance – with a highly biased result may lead to overconfidence). Cosmides and Tooby (1996) demonstrate that reasoning better emulates probability theory when people are obliged to consider the frequency of events.

Although people may not base their beliefs on a precise theory of probabilities, studies have shown that people’s responses to inference problems are related to the probabilities associated with the propositions used.

Oaksford, Chater and Larkin (2000) presented conditional inference tasks following scenarios which varied the probability of different propositions. Their inference tasks provided a conditional relationship (e.g. ‘If a card has an S on the front, then it has a 5 on the back’), a proposition (e.g. ‘This card has an S on the front’) and conclusion (e.g. ‘Therefore this card has a 5 on the back’ – in this case, based on MP inference – see

§1.1). Oaksford et al. (2000) demonstrated that the probability that a proposition is true has an effect on people's willingness to endorse inferences using the proposition. The probability of the propositions was initially manipulated explicitly using scenarios based on a machine that printed cards with different shapes of different colours on them. The scenarios suggested that the probability of a proposition like 'the card is red' is high with statements like, 'out of every 60 cards, 40 are red'. They suggested that the probability of a proposition stating that any card was a given colour is low with a statement like, 'roughly 12 out of every 60 cards are of each colour'. Oaksford et al. (2000) found a strong influence of the probabilities given in these explicit scenarios on participants' ratings of the conditional inferences. They also used implicit variation of probabilities by presenting conditional rules made up of pretested propositions that had low probability (e.g. 'an animal is a chipmunk') or high probability (e.g. 'an item of furniture is heavy'). These implicit manipulations also affected participants' willingness to endorse inferences using the propositions given although the results were not as strong as where probabilities were explicitly provided.

Subsequent studies have also shown that people use probability information to provide estimates of the likelihood of propositions. Evans, Handley and Over (2003) asked participants to judge the likelihood that 'if  $p$  then  $q$ ' was true given frequency information about instances where  $p$  and  $q$  were true, *not*  $p$  and  $q$  were true,  $p$  and *not*  $q$  were true and *not*  $p$  and *not*  $q$  were true. They found that the estimated likelihood of the conditional statement was related to either the conditional probability ( $P(q|p)$ ) or the probability of the conjunction ( $P(p \text{ and } q)$ , sometimes written  $P(p \cap q)$ ). Similar results, showing the importance of the conditional probability  $P(q|p)$  to people's evaluation of the conditional 'if  $p$  then  $q$ ', were reported by Over and Evans (2003) and Oberauer and Wilhelm (2003). These results demonstrate the psychological relevance of the conditional probability hypothesis. This hypothesis arises in the philosophical literature and states that the probability that a statement of the form 'if  $p$  then  $q$ ' is true is equivalent to the conditional probability  $P(q|p)$  (Ramsey, 1931; Adams, 1987; Edgington, 1995).

Ohm and Thompson (2006) explored this conditional probability hypothesis. They provided participants with scenarios that used conditional inducements (threats and promises) and conditional advice (warnings and tips). For each scenario, participants



were asked to estimate the probability that each of four possibilities was true ( $p$  and  $q$ ;  $not\ p$  and  $q$ ;  $p$  and  $not\ q$ ;  $not\ p$  and  $not\ q$ ). For example, a scenario that provided conditional inducement (a warning) was “John is not attending all of his classes at university. While talking to his roommate, he is told that ‘If you skip classes, you will fail your courses’”. Conditional probabilities  $P(q|p)$  and  $P(q|not-p)$  were derived from participants’ responses and compared to participants’ evaluation of the conditional rules. As predicted by the conditional probability hypothesis, participants’ evaluation of conditional inferences varied as a function of  $P(q|p)$ . Ohm and Thompson (2006) also looked at how pragmatic factors in the conditional statements led to different estimates of  $P(p)$  suggesting the different likelihoods that participants attributed to the conditionals’ ability to change behaviour. They found a better correlation between  $P(q|not-p)$  and these differences than between  $P(q|p)$  and these differences.

Research looking at people’s naïve application of probability theory has shown a range of biases as with the naïve application of formal logic. This may suggest that people are not able to use probability information and estimates of likelihood in reasoning. However, a range of studies have shown that people do appear to use evidence about the probability of propositions in responding to deductive reasoning tasks. It is possible that this use of likelihood influences the negation effects observed in tasks that are notionally intended to test formal logic. This can be tested by looking at whether different types of negation, that give rise to different estimates of likelihood, lead to different responses in reasoning tasks. Before looking directly at the use of different types of negation in conditional reasoning tasks, this chapter will first explore how the type of negation used in a proposition influences participants’ estimates of the likelihood of that proposition.

### **2.1.1. Probability Rating Tasks**

The experiments reported in this chapter each provided participants with a series of scenarios followed by probability rating task questions (Oaksford, Chater & Grainger, 1999). Probability rating tasks present a description of a group of people and ask how many of these people meet the circumstances of the given proposition. The response provides information about the perceived probability that the given proposition is true

for a person in the group described. These experiments looked at whether responses vary when propositions use different types of negation.

In particular, the present experiments considered whether people respond differently depending on whether the word “not” or an affix like “un-” is used to negate a proposition. As discussed below, different types of negation can have different implications (e.g. Horn, 1989). However, the reasoning literature on negation has not considered these implications. Propositions that use forms of negation with different implications would be expected to have different probabilities (for example a proposition using “not” negation with large scope might be expected to have a larger probability than the same proposition using narrow scope affixal negation). If such propositions were shown to have a differential effect on responses to reasoning problems, this would undermine models of reasoning which assume that all types of negation have a single logical function.

## **2.2. Different Types of Negation**

The previous chapter observed that there were many different ways of expressing negation which could give rise to different understandings and emphasises of the scope of negation and what the negation means. Different forms of negation can imply different levels of strength and different scope. Although negation can be seen as having the single, simple, logical implication of denying a proposition in classical logic, other forms of logic allow it to fulfil different logical roles such as Aristotle’s contrary and contradictory operators (Horn, 1989).

The most obvious use of negation in language is the application of a clear, separate negative operator, like ‘not’, before the proposition or term being negated. Words like ‘no’, ‘never’, ‘none’ and ‘nobody’ act as clear negative operators in appropriate contexts. The present chapter will look at whether negation using a clear negative operator ‘not’ in propositions is interpreted differently to two other forms of negation: affixal and implicit negation. Affixal negation uses prefixes, like ‘un-’, or suffixes, like ‘-less’ to negate the meaning of a stem word. Implicit negation uses an affirmative statement to negate a proposition which might otherwise have been understood to be true.

### 2.2.1. Affixal Negation

Affixal negation uses prefixes or suffixes to negate a base word. Common prefixes in modern English include “un-”, “iN-” (which includes prefixes “in-”, “il-”, “ir-” and “im-” depending on the first letter of the base) and “non-” (Mazzon, 2004). Where statements using ‘not’ tend to create a contradictory interpretation of the negation, these affixal operators tend to create a more contrary interpretation (Zimmer, 1964). For example, the statement ‘I am unhappy’, is normally interpreted as analogous to ‘I am sad’ (the opposite of ‘I am happy’). However, ‘I am not happy’ could indicate that the speaker has one of many emotional states besides happiness (such as ‘I am angry’ or ‘I am disappointed’).

Jespersen (1917) hypothesised that only words with emotionally positive implications (e.g. *decisive*, *happy*) could be subject to negative affixes and the negative compounds created were emotionally negative (e.g. *indecisive*, *unhappy*). Compounds like “unsad” or “inhesitant” do not routinely occur in English providing superficial support for this proposal. However, a thorough analysis of the corpus of affixally negated words in several languages found that Jespersen’s (1917) hypothesis would not allow many often observed uses of affixal negation (Zimmer, 1964). In English, Zimmer (1964) found that most base words that were affixally negated using “un-” are neither emotionally positive or negative (e.g. *uncaught*, *uneaten*). He also found that “un-” could be used to negate bases with negative connotations (e.g. *unobjectionable*). Going further, he showed that the “un-” prefix can be used to negate any base words with deverbal suffixes (“-able”, “-ed”, “-ing”). Zimmer (1964) also analysed corpora of affixally negated words in other languages and found that they can apply affixes with even fewer restrictions. For example, the use of the Russian prefix “ne-” is not limited or biased at all by the emotional implications of the base. Therefore, each affix may have different applications and implications and we must, therefore, look at each form of affixal negation separately.

In English, compounds using the prefix “iN-” are normally understood to be contrary opposites of their base words. For example, ‘indiscrete’, ‘illogical’, ‘irretrievable’ and ‘immaterial’ are interpreted as opposites of ‘discrete’, ‘logical’, ‘retrievable’ and ‘material’. There is a relatively narrow corpus of compound words which incorporate the affix “iN-” (Zimmer, 1964, found around 300 adjectives in his study of English

corpora). This appears to be because the prefix has limited productivity in modern English, it is no longer used to create novel compounds of negated bases (Zimmer, 1964). The prefix was most productive during the late Middle English and early Modern English periods when many new compounds were generated (Mazzon, 2004). New applications of the “iN-” prefix only occur when an author wants to sound formal (Zimmer, 1964). The prefix “iN-” is generally recognised as a negative operator that modifies a base word and novel formulations can be understood even if they sound odd. However, most compounds using “iN-” in modern English appear to have their own lexical entry as part of people’s learned vocabulary (Mazzon, 2004). There are some “iN-” compounds, like ‘impeccable’, where the lexicalised compound is the more widely recognised base (while ‘peccable’ may still be in some dictionaries it is not as well recognised as the derived compound, Zimmer, 1964). As they are highly lexicalised, words affixally negated using “iN-” can develop specific meanings in their own right. They are therefore typically interpreted as having highly contrary implications compared to other forms of affixal negation which are less lexicalised (Horn, 1989).

In Modern English, the prefix “un-” is observed much more frequently than “iN-”. Zimmer (1964) found around 2,700 adjectives derived using “un-” in his review of English corpora. Mazzon (2004) describes “un-” as highly productive in Modern English and is frequently used to create novel compounds. Zimmer (1964) also found “un-” to be highly productive. He pointed to readily understandable compounds which are not in dictionaries (like *uncommemorated*, *uninitialled* and *unconstructive*). Mazzon (2004) showed that “un-” could be prefixed to most types of adjectives and Zimmer (1964) showed that it could also be applied to many other types of word. However, the application of “un-” is not unrestricted. It is unusual for compounds to be formed using “un-” when there is already a lexicalised “iN-” compound available (although such compounds are likely to be readily understood, Zimmer, 1964; Mazzon, 2004). Where “un-” has been found in written corpora compounded with a base that already has an “iN-” derivation, it is typically where the “un-” compound yields a useful different understanding (see below for examples). These “un-” compounds normally have a less contrary and more contradictory interpretation to their “iN-” equivalents (Horn, 1989). A similar restriction to the application of “un-” applies to adjective words where there

are short words available in the lexicon which provide contrary opposition. For example, *ungood* is readily understood by English speakers, but the uncompounded alternative, *bad*, is normally favoured (Zimmer, 1964). The prefix “un-” can be used to suggest complementary and contrary understandings (Mazzon, 2004). Horn (1989) therefore suggested that “un-” has a function between not (which typically yields contradictory interpretations) and “iN-” (which provides a highly contrary form of negation).

The prefix “non-” appears to be increasingly used in Modern English (Mazzon, 2004). It is readily productive and often used to generate novel compounds (Zimmer, 1964). It has few restrictions (although it is less often applied to simple adjectives with well used short, simple antonyms, like ‘short’ where a person would tend to use ‘long’ rather than ‘nonshort’, Zimmer, 1964). In contrast with “un-” and “iN-”, “non-” can even be used with already affixally negated bases to create new compounds like ‘noninfinite’. Compounds using “non-” are interpreted as negating the descriptive meaning of their base rather than the evaluative meaning (Zimmer, 1964). For example, compare the different understandings of ‘unnatural’ (which is something that does not have the ‘value’ of natural) and ‘nonnatural’ (which is simply something that does not occur naturally). This provides a more contradictory understanding of “non-” negation than “un-” or “iN-” negation, an understanding closer to “not” negation (Horn, 1989). However, unlike “not” negation, “non-” negation has a very narrow scope – applying just to the base word. This means that “non-” compounds are rarely ambiguous in the way that “not” negation can be (Zimmer, 1964).

Other types of affixal negation in the English language (including prefixes “dis-” and “a(n)-” and suffix “-less”) only have very limited productivity (Mazon, 2004). The experiments in this chapter will, therefore, apply the types of affixal negation described above. These provide a spectrum from low productivity and highly contrary interpretation (“iN-”) through fairly productive and intermediate contrary and contradictory interpretation (“un-”) to high productivity, more contradictory interpretation (“non-”). Looking at some of the base words to which all three affixes are readily applied demonstrates the different understandings that they create in their compounds. For example, “human” can be negated to create *inhuman*, *unhuman*, *nonhuman* and “religious” generates *irreligious*, *unreligious*, *nonreligious*. The words

in these sets have more contrary meanings in their “iN-” form, intermediate interpretations in their “un-” form and more contradictory interpretations in their “non-” form (Horn, 1989). Looking at how people process different forms of negation with different implications may help to show how people interpret the implications of negation when reasoning. If responses vary when using different forms of negation, this would suggest probabilistic effects that are easier for some models of reasoning to explain than others.

### **2.2.2. Implicit Negation**

One of negation’s core roles is to reject propositions that might otherwise be assumed true. This can obviously be observed when using a clear negative operator in a sentence like ‘David is *not* in Manchester’ (which rejects any notion of David being in Manchester) or ‘Patricia is *unwell*’ (rejecting the proposition that ‘Patricia is well’). Such application of negative operators is described as ‘explicit negation’ (Horn, 1989). However, propositions can also be rejected without the application of a clear negative operator. An affirmative statement that contradicts a proposition implicitly negates the proposition. For example, the sentences ‘David is in Birmingham’ or ‘Patricia has the flu’ used to reject the propositions in the previous examples. The word ‘implicit’ has been used to describe several negation phenomena in studies of language.

Tottie (1982) conducted analyses of spoken and written English corpora and proposed a categorisation of the forms of negation observed. She divided occurrences of negation into the categories of ‘rejection’ (in spoken language, this would typically take the form of ‘No’ and then a negative statement) and ‘denial’ (which is used to state that a supposition is false). The ‘denial’ category was subdivided into ‘explicit denial’ and ‘implicit denial’. Tottie’s (1982) subcategory of ‘implicit denial’ referred to a statement that negated a proposition without explicitly referencing the negated proposition.

Alternative attempts to categorise implicit negation have focused on affirmative statements which are analogous to negative expressions but which do not use obvious negative operators. For example, the use of ‘forbid’ for ‘not allow’ or ‘absent’ for ‘not present’. Klima (1964) sought to provide criteria for the identification of negative sentences. He proposed that negative elements were those that can co-occur with *any* in a clause or with *either* when the negative is the latter of two clauses conjoined with

*and*). For example, these criteria define ‘none’ as negative but not ‘all’ (‘none of the people did any work’ is acceptable while ‘all of the people did any work’ is not). They also categorise words such as *few*, *seldom* and *little* as negative but not those like *absent* or *forbid*. Klima (1964) therefore suggested that these latter phrases, which did not meet his criteria for negative elements but could be used to express negative sentiments, were examples of ‘inherent negation’ (Horn, 1989, also used this phrase as a label for implicit negation). Clark (1974) discussed the suggested distinction between those phrases identified as negative by Klima’s (1964) criteria and implicit negation. Clark (1974) argued that the former asserted that a supposition was false (e.g. ‘Pete didn’t remember’ denies the supposition that Pete remembered) while the latter affirmed a negative supposition (e.g. ‘Pete forgot’ a statement which affirms the negative supposition that Pete did not remember).

Context appears to be very important in how we use and understand implicit negation. Clark (1974) looked at a series of studies that had compared participants’ processing time when completing tasks using explicit negation, implicit negation or affirmative propositions. He found that responses to tasks using implicit negation were quicker than where the negation was explicit and slower than responses to tasks using affirmative propositions. Clark (1974) related this finding to Wason’s (1965) ‘contexts of plausible denial’ theory. Wason (1965) hypothesised that people understand negative propositions by transforming them into affirmative propositions to understand them. He suggested that such transformations would be made more quickly where there was reason to believe that the proposition being negated might be true (i.e., the negated proposition existed within a context of plausible denial). For example, Wason (1965) suggests a statement that “the whale is not a fish” is more easily understood than “the whale is not a bird”. This is because the former statement contradicts a proposition that someone who did not know better might assume whereas the latter statement contradicts a proposition that no one is likely to believe.

Some explanations have been suggested for this context effect. Wason (1965) sought to explain plausible denial as a property of sets (the negated proposition is an exception or low incidence case). Clark (1974) suggests more simply that negative sentences combine a supposition and a negation and are easier to process if the context makes the supposition plausible. Clark (1974) uses the example that if a sports team was

predominantly male the supposition that a given team member is male is plausible so combining that supposition with a negation (e.g. ‘player number ten isn’t male’) is easier to process. The opposite supposition, that a given team member is female, is not plausible in this context and therefore it takes longer to process that supposition combined with a negation (e.g. ‘player number one isn’t female’).

Horn (1989) builds on this, suggesting that implicitly negative statements are only used where the supposition being negated is explicitly suggested in context (and therefore that in processing the statement the supposition being negated does not need to be constructed as it might be if explicit negation is used). Giora, Fein, Metuki and Stern (2010) found that using explicit negation did not remove or reduce the negated concept in a person’s mental representation but may enhance information within the scope of the supposition allowing a metaphorical interpretation. For example, the statement ‘I am not your maid’ in response to being asked to clear something up does not suppress the notion of ‘maid’. Rather, it uses the concept to create a metaphoric interpretation in a manner that a positive statement (e.g. ‘I am your maid’) does not. Giora et al. (2010) also observed this metaphor inducing effect with implicit negation (for example, the statement ‘What am I, your secretary?’ implies negation and was interpreted metaphorically rather than literally by most of their respondents).

Statements using implicit negation do not generally display the logical ambiguity seen in “not” negation which often relates to the perceived scope of the negative operator. Clark (1974) suggested that his review of prior studies supported the ‘scope of negation’ hypothesis. This hypothesis says that the wider the scope of negation in a proposition the longer the processing time required to deal with the proposition. For example, a statement with wide scope, like ‘A isn’t present’, would take longer to process than one with narrow scope, like implicit near equivalent ‘A is absent’. The hypothesis is based on the idea that the wider the scope of a proposition, the more difficult it is to comprehend. Based on this hypothesis, Clark (1974) suggested that, because implicit negation asserts an affirmative proposition with narrower scope than an equivalent statement using explicit negation, processing times are lower for implicit negation than “not” negation.



Previous studies of human reasoning have used implicit negation extensively, particularly tests using Wason's (1968) selection task. For example, Evans and Lynch (1973) provided participants with cards marked 'S', '9', 'G' and '4'. They asked them which they should turn over to verify the rule 'If there is an S on one side then there is a 9 on the other side'. In this context, the 'G' card represents 'not-S' and the '4' card 'not-9'. These cards, therefore, constitute implicit negations of the propositions referred to in the rule. Evans and Lynch (1973) observed that participants selected cards labelled with the propositions in the rule rather than the logically correct responses. This suggests a matching bias in line with other selection task experiments (Evans, 1998) but Jackson and Griggs (1990) have claimed that subsequent studies show that other factors may help to reduce matching bias when cards use explicitly negative statements. For example, Cheng and Holyoak (1985) presented the selection task and compared responses to rules like "If one is to take action 'A', one must first have fulfilled prerequisite 'P'" and cards which used explicit negation (e.g. 'has taken action A', 'has not taken action A', 'has fulfilled prerequisite P' and 'has not fulfilled prerequisite P') to rules like "If a card has an 'A' on one side, then it must have a '4' on the other side" and cards using implicit negation ('A', 'B', '4', '7'). They found that participants gave responses more in line with the logically correct answer when explicit negations were used because these materials were more deontic (that is, used rules with a stronger sense of obligation).

Because implicit negation has been used widely in the reasoning literature, the first experiment will look at its effect on people's estimates of the degree to which they should draw a particular conclusion alongside other forms of negation.

### **2.3. Experiment 1**

The experiments in this chapter are aiming to explore whether using different types of negation in propositions gives rise to different estimates of the likelihood of the proposition. As previous research has suggested, different likelihood estimates may influence responses to logical reasoning tasks using negation. This first experiment (previously reported by Vance, unpublished Masters dissertation, 2011) is intended to provide an initial test of whether different types of negation leads to different estimates of the likelihood of propositions.

This experiment will look at people's responses to three probability rating tasks (Oaksford, Chater & Grainger, 1999) for each scenario. The rating tasks will ask how many of 100 people in the scenario are truly described by a given proposition. In one question for each scenario, the proposition will be affirmative. In another, it will be "not" negated. The third question for each scenario will use an alternative form of negation, either affixal negation (using one of three prefixes, "non-", "un-" or "iN-") or implicit negation.

The implicit negation used in previous reasoning studies has taken a range of different forms. For the current experiment, the implicit negation propositions used terms which were contrary opposites of the affirmative proposition. These terms were selected from those used by Bianchi, Savardi and Kubovy (2011) who developed a model to describe the perception of opposite terms within a multidimensional conceptual space. Bianchi et al. (2011) presented participants with a series of concepts and their opposites and asked them to consider spatial scales representing the full range of spatial experiences between those opposites. For example, the provided opposites like 'large' and 'small' at either end of a horizontal scale. People were asked to use the scale to indicate where on the scale the boundary between the opposites would occur (e.g. the line between 'large' and 'small'). They were then asked to mark the space on the scale that was neither one concept nor its opposite (e.g. the part of the scale that would represent neither 'large' nor 'small'). The proportions of the spatial scales identified by Bianchi et al. (2011) may relate to the probability estimates made by participants in the present experiment about the concepts and their opposites.

For each of the 100 hypothetical people described in one of the scenarios in this experiment, the affirmative version of the proposition could either be true or not true. Therefore, the number of hypothetical people in a scenario for which the proposition is either true or not would, in reality, total 100. Judgements made by a single participant asked both for how many of the hypothetical people a proposition is true and for how many it is not true are therefore not independent. A participant that made a high judgement for the likelihood of a given affirmative proposition would be expected to make a low judgement for the "not" negated proposition related to the same scenario. For example, if someone thought 90 of the people described were 'large', you would expect them to think 10 were 'not large'. Whereas, if that person thought the probability

of the affirmative proposition was low (e.g. they thought 10 would be ‘large’) then you would expect them to think the probability of the “not” negation proposition would be high (they would think 90 would be ‘not large’). To compare the effects of different types of negation within this experiment’s within-participants design, we will, therefore, compare participants’ additive scores for their responses to each question using negations. These additive scores are simply the response to the negation question added to the response to the relevant affirmative question with 100 subtracted (e.g. if someone thought 54 people would be ‘large’ and 48 ‘not large’ the additive score for ‘not large’ would be  $54 + 48 - 100 = 2$ ). Using additive scores mitigates the effect of differences in perceptions of the likelihood of the affirmative proposition in the subsequent analysis.

Macchi, Oshero and Kranz (1999) also looked at people’s estimates of the probability of events which could be divided into mutually exclusive sub-events. They described estimates of the probability as ‘superadditive’ if the sum of the perceived probabilities of two mutually exclusive sub-events was less than the perceived probability of the event combining them. They found such judgements were more likely where participants had limited knowledge about the events considered. Similarly, they described estimates as ‘subadditive’ where the sum of probabilities given for sub-events exceeded the probability of the overall event. Macchi et al. (1999) suggested that people may make such interpretations in contexts where they have extensive knowledge of the events. In this experiment, a negative additive scores would indicate superadditive results and positive additive scores would indicate subadditive results.

The placement of the word ‘not’ in “not” negation may affect its understanding – for example, it can affect the scope of negation (Horn, 1989). Statements like ‘... how many would you not expect to be ....’ would appear to be more natural English language statement than the form ‘... how many would you expect to be not ....’ where the scope of negation is logically clearer – placing the negative function immediately before the item to be negated. As the present study is interested in how the uses of negation in natural language influence people’s responses to logical reasoning tasks, this experiment will also explore whether people respond differently to these two forms of “not” negation. The results will inform whether it is reasonable to use more naturalistic forms of “not” negation in further experiments.

### **2.3.1. Predictions**

This experiment will present scenarios that provide participants with very limited information on which to make a judgement on the likelihood of different propositions. Macchi et al. (1999) found predominantly superadditive responses in similar circumstances. This suggests that all additive scores will be negative.

The key question in this chapter is whether different types of negation lead to different estimates of the likelihood of propositions using negation. The alternative forms of negation used (affixal and implicit negation) both normally have more contrary interpretations than “not” negation (Horn, 1989). More contrary statements of negation tend to have narrower interpretations than those provided by related contradictory statements. We would, therefore, expect estimates of a proposition’s likelihood to be lower where the proposition uses an alternative form of negation (affixal negation with “iN-”, “un-” and “non-”, and implicit negation) than where the proposition uses “not” negation. This would be observed as a more negative additive score.

Within the three different types of affixal negation used, “iN-” typically has a more contrary interpretation than “un-” which, in turn, normally takes a more contrary interpretation than “non-” (Zimmer, 1964; Mazzon, 2004). This suggests the prediction that propositions using “iN-” will give rise to lower (more negative) additive scores than propositions using “un-” which in turn will be lower than those involving “non-”.

### **2.3.2. Method**

#### **2.3.2.1. Design**

A within-participant design was used with additive scores calculated as described above as the dependent variable. The independent variable was type of negation which had five levels (“not” negation, affixal negation with “iN-”, “un-” and “non-”, and implicit negation). An additional between participants variable, position of ‘not’ in questions using “not” negation was also included to look at whether this mattered to people’s interpretation of the task.

#### 2.3.2.2. Participants

An opportunity sample of 20 participants (12 female) aged between 24 and 63 (median age 40) took part in the experiment. Participants were all fluent in English (18 had English as their first language) and participated in groups of between one and three (each participant worked silently on their question booklet at their own pace).

#### 2.3.2.3. Materials

Eight scenarios for each of four alternative negation conditions (Thirty-two scenarios in total) were prepared. The scenarios each consisted of two statements. The first statement in each scenario briefly set out something about the circumstances of 100 hypothetical people. The second statement was intended to indicate that there would be differences between the hypothetical people in relation to the propositions that participants would be asked about. The second statements were intended to be ambiguous and not indicate whether the affirmative or negative versions of the proposition would be true for greater or fewer numbers of the hypothetical people. For example:

*100 people are asked to change their shift patterns at work.*

*The boss knows that their workers will have different levels of flexibility in being able to change shifts.*

For each scenario, three questions were provided: for the affirmative and two relevant negation conditions (“not” negation and whichever of the alternative negation the scenario used, either “iN-”, “un-”, “non-” or implicit negation). For example, the following questions were provided for the scenario above (which used “iN-” negation as its alternative negation):

*Out of the 100 people, how many would you expect to be flexible?*

*Out of the 100 people, how many would you not expect to be flexible?*

*Out of the 100 people, how many would you expect to be inflexible?*

Examples of questions from other scenarios using the other forms of alternative negation are as follows (with “un-”, “non-” and implicit negation respectively):

*Out of the 100 people, how many would you expect to be fit?*  
*Out of the 100 people, how many would you not expect to be fit?*  
*Out of the 100 people, how many would you expect to be unfit?*

*Out of the 100 people, how many would you expect to be traditional?*  
*Out of the 100 people, how many would you not expect to be traditional?*  
*Out of the 100 people, how many would you expect to be nontraditional?*

*Out of the 100 people, how many would you expect to be large?*  
*Out of the 100 people, how many would you not expect to be large?*  
*Out of the 100 people, how many would you expect to be small?*

Response booklets provided instructions on the first page followed by 96 question pages. Each question page presented a scenario followed by one of the three questions related to it (so each scenario was provided three times within each booklet, once with each question). There were 96 question pages in total, 32 with an affirmative polarity question, 32 with a “not” negated question and 8 for each of the four alternative negation conditions. Each booklet had the questions pages in random order.

To test whether the placement of the word ‘not’ in “not” negation affected people’s probability estimates based on propositions using it, two different forms of the “not” negated questions were used. Half of the booklets used “not” negation questions of the form ‘Out of the 100 people, how many would you not expect to be ...’. The other half used the form ‘Out of the 100 people, how many would you expect to be not ...’.

Full information about the materials used is provided in appendix 1 (see §A1.1).

#### 2.3.2.4. Procedure

Prospective participants read an information sheet and, if they agreed to participate, signed a consent form and completed a demographic questionnaire. They were then given the test booklet and the experimenter read the instructions (which were also printed on the front of the booklet). The experimenter responded to any questions by referencing these instructions. The participant was then invited to open the booklet and complete each question in turn. They were given as much time as needed to complete

every question. Once finished, participants were able to ask any questions that they had about the study.

### 2.3.3. Results

Participants were asked to answer 96 questions each. However, six participants provided no answers to fourteen questions between them. The mean response to the given question provided by other participants was substituted for these missing responses in the following analysis. The responses were then converted to additive scores as follows:

$$\begin{aligned} \text{Additive score} = & \quad [response\ to\ question\ using\ negation\ in\ proposition] \\ & + [response\ to\ question\ with\ affirmative\ proposition] \\ & - 100 \end{aligned}$$

To look at whether the position of ‘not’ in the questions using “not” negation had any effect, half the participants saw questions using “not” negation with the form ‘be not’ and the other half saw the form ‘not expect’. An independent samples t-test compared the average scores for “not” negation between these groups and found no significant effect ( $t(18) = -0.024$ ,  $p = .981$ ). It, therefore, appears that this variable did not affect responses. Therefore, responses from both groups to “not” negation questions were combined for subsequent analysis.

#### 2.3.3.1. Overall Differences between Different Types of Negation

Table 2.1 shows the overall average score for each negation condition. As predicted, they are all negative. This is consistent with Macchi et al.’s (1999) previous finding that people make superadditive estimates when they have very limited information about the scenarios considered.

*Table 2.1 – Summary of responses to chapter 2, experiment 1*

Type of Negation	“not” negation	<i>iN</i> - affixal negation	<i>un</i> - affixal negation	<i>non</i> - affixal negation	Implicit negation
Mean Score	-3.0 (6.0)	-4.6 (7.3)	-2.0 (10.9)	-4.4 (5.1)	-21.7 (14.0)

Mean additive score (standard deviation) to 1 decimal place

Participants' mean scores for each type of negation were compared using a one-way repeated measures ANOVA. This found a significant effect ( $F(4,76) = 20.480, p < .001$ ) of substantial size ( $\eta_p^2 = .519$ ). A series of paired samples t-tests with Bonferroni adjusted significant levels ( $p \leq .005$ ) compared each type of negation with each other type of negation to understand the nature of this effect. Significant differences were only found between implicit negation and each of the other forms of negation ("not" negation:  $t(19) = 5.857, p < .001$ ); "iN-" negation:  $t(19) = 6.141, p < .001$ ; "un-" negation:  $t(19) = 5.5642, p < .001$ ; and "non-" negation:  $t(19) = 5.417, p < .001$ ). Therefore the effect of type of negation appears to represent a difference between the implicit negation condition and all of the other negation conditions which yielded similar results.

### 2.3.3.2. Affixal Negation

Each type of affixal negation was tested using a different set of eight scenarios. Therefore, it is interesting to look at each set of eight scenarios separately to consider how responses when each type of affixal negation was used varied from responses when "not" negation was used with the same set of scenarios. This analysis, therefore, breaks down responses to tasks where "not" negation is used into four groups (one for each set of scenarios in which a different type of alternative negation used – three different affixes and implicit negation). The additive scores for "not" negated questions referred to below, therefore, vary slightly from the overall average score given above in table 2.1.

For the scenarios which used "iN-", the mean additive score for the "not" negation questions was -3.1 (standard deviation 10.5), a little higher than the mean scores for "iN-" negation questions. A one-tailed paired-samples t-test found that this difference in average scores was not significant ( $t(19) = 0.723, p = .239$ ). Looking at each of the eight scenarios which used "iN-" negation in turn, mean additive scores for "not" negation were higher than scores for "iN-" negation for five scenarios. However, when compared using one-tailed paired samples t-tests, these differences only approached significance for one scenario ( $t(19) = 1.572, p = .067$ ). This scenario used the terms 'not significant' and 'insignificant' in its "not" negated and affixally negated conditions respectively. It is possible that participant familiarity with statistical methods provided a better understanding of the difference between these terms.



The scenarios using “un-” negation provided a mean additive score for the “not” negated condition of -3.0 (standard deviation 10.9). The mean additive score for “un-” negated questioned is therefore slightly higher than for the equivalent “not” negated questions, but this difference was not significant ( $t(19) = -0.551$ ,  $p=.588$ ). Within each scenario’s scores for “not” negation were higher than for “un-” negation in four scenarios but this order was reversed in the other four. Paired-samples t-tests found no significant, difference between scores for *not-* negation and “un-” negation for any scenario.

Where scenarios used “non-” in their alternative negation condition, the mean additive score for responses to the “not” negated questions was -3.8 (standard deviation 6.7). Therefore “non-” negation scores are lower than “not” negation scores, but this is not significant based on a one-tailed paired-samples t-test ( $t(19) = 0.422$ ,  $p=.339$ ). Across the eight scenarios, four yielded higher additive scores for “not” negation and four for “non-” negation. No significant differences between “not” negation and “non-” negation scores were observed for any of the eight scenarios in paired-samples t-tests.

#### 2.3.3.3. Implicit Negation

Like the individual forms of affixal negation, implicit negation was used in only eight scenarios. For just these scenarios, responses to “not” negation questions had a mean additive score of -2.2 (standard deviation 8.3). Therefore, average implicit negation scores are lower than “not” negation scores. A one-tailed paired-samples t-test demonstrated that this difference was highly significant ( $t(19) = 6.164$ ,  $p<.001$ ).

Seven of the eight scenarios which used implicit negation showed this predicted order of scores (with responses to questions with “not” negation being higher on average than responses to questions with implicit negation). One-tailed paired-samples t-tests showed that the differences in average scores were significant for three scenarios. These were: those using ‘not fat’ and ‘slender’ ( $t(19) = 5.097$ ,  $p<.001$ ); those using ‘not large’ and ‘small’ ( $t(19) = 5.913$ ,  $p<.001$ ); and those using ‘not top’ and ‘bottom’ ( $t(19) = 9.652$ ,  $p<.001$ ).

As explained above, the affirmative stem words and implicitly negating opposite phrases used in this experiment were drawn from amongst the words used for

affirmative concepts and their spatial opposites by Bianchi et al. (2011). Table 2.2 shows the proportions of the spatial space that each concept occupied for participants in Bianchi et al.'s (2011) task and the proportion suggested by the average responses to the present experiment. Overall, the proportions established by Bianchi et al. (2011) were not significantly correlated with the proportions based on responses to the current experiment ( $r = .255$ ,  $p = .341$ ). However, just the eight affirmative concepts had a correlation which approached significance ( $r = .676$ ,  $p = .066$ ) and the eight opposite concepts showed a significant correlation ( $r = .763$ ,  $p = .028$ ) between the results of this study and that of Bianchi et al. (2011). Bianchi et al.'s (2011) task derived estimates of proportion from group work between participants so participant level data cannot be statistically compared with the results of the present experiment. However, it is apparent from the data reported in table 2.2 that responses differed. Only three of the sixteen concepts considered show estimates of proportion from Bianchi et al.'s (2011) study within 0.1 of the estimate derived from the results of the current experiment.

*Table 2.2 – Comparing proportion estimates from Bianchi et al. (2011) and present experiment*

Concept			Opposite Concept		
Label	Bianchi et al. estimate of proportion	Present experiment estimate of proportion	Label	Bianchi et al. estimate of proportion	Present experiment estimate of proportion
moving	.82	.64	still	.15	.34
above	.47	.52	below	.33	.57
fat	.46	.33	slender	.22	.37
open	.76	.45	closed	.10	.49
large	.44	.27	small	.21	.32
top	.31	.11	bottom	.10	.12
in front	.56	.14	behind	.37	.76
deep	.39	.34	shallow	.18	.47

#### **2.3.4. Discussion**

This experiment sought to test the hypothesis that different types of negation would have different effects on estimates of a proposition's probability. The experiment found an overall effect of type of negation as expected. However, this effect can be attributed

to the difference between implicit negation and all the other types of negation. The predicted differences between “not” negation and the forms of affixal negation tested were not observed.

It was anticipated that estimates of the probability of propositions using a highly contrary form of negation would be lower than for propositions using “not” negation. The result that implicit negation led to lower scores than “not” negation in probability rating tasks is consistent with that hypothesis. By showing that people perceive different probabilities to statements using explicit and implicit negation, this result supports a probabilistic rationale for previous results showing differences in responses to reasoning tasks when implicit and explicit negation are used (e.g. Jackson and Griggs, 1990; Cheng and Holyoak, 1985).

It was not anticipated that all forms of affixal negation would show significantly different response levels to implicit negation. Such a difference for more contradictory forms of negation (like “non-”) was expected. However, a much smaller difference – or no difference – was predicted where the negation was more contrary (like “iN-”) where its interpretation was expected to be closer to that of implicit negation. However, different word stems were used to test each different form of alternative negation making direct comparison problematic. It was anticipated that, overall, affixal negation would provide a more contrary implication than implicit negation. Therefore, the differences observed may not be inconsistent with the view that different forms of negation with different levels of contradictory or contrary implication would have different effects.

It was anticipated that probability ratings would be lower when using different forms of affixal negation than when using “not” negation. However, no significant overall differences between any type of affixal negation and “not” negation were found. While the results did not support the hypothesis overall, they did not contradict it either. Some results suggest the hypothesis is plausible, even if the present experiment is not sufficiently sensitive to detect what may be small effects. Mean scores were lower for two of the three types of affixal negation (*iN*- and “non-”) than for “not” negation. More than half of the scenarios showed average scores in the direction proposed by the hypothesis. The only scenario where the difference between affixal and “not” negations

approached significance showed lower scores in response to affixal negation than “not” negation as expected. This result was also observed with “iN-” negation. This form of affixal negation is generally interpreted to be a more contrary form than “not” or “non-” negation (Zimmer, 1964; Mazzon, 2004). It may be that any effects would be more easily observable with this form of affixal negation (i.e., “iN-”) than with the other forms tested. Demonstrating significant differences between the probability ratings of propositions using different forms of negation could have implications for our understanding of reasoning processes. If different forms of negation have different probabilistic implications, then probabilistic models of reasoning (e.g. Oaksford et al., 2000) would be able to account for these differences better than other models that treat negation as a simple operator.

If there are systematic differences between responses using affixal negation and “not” negation, they are likely to be small and a more sensitive methodology may be required to observe them. It is also not possible to compare the results for different forms of affixal negation in the current experiment because each form of alternative negation used different scenarios. Scenario effects are quite likely, potentially obscuring the effects of different types of negation. This interpretation is consistent with the reduction in the size of negation effects seen when the results were analysed by scenario. The responses previously obtained by Bianchi et al. (2011) which showed people would distinguish between a given proposition and its implicitly negating opposite were not seen here which also suggests that the method used may not be sufficiently sensitive to observe established effects.

## **2.4. Experiment 2**

This experiment is intended to provide a further test of the hypothesis that people will make lower probability rating task judgements when affixal negation is used than when “not” negation is used. It makes several changes to the previous method intended to make the tasks more sensitive to detect small effects.

The within-participants design of the previous experiment meant that every participant was answering questions for each scenario covering each of the negation conditions. Although we would expect participants to interpret affixal negation differently to “not” negation, participants may have noted the similarity in the questions and sought to

answer them consistently. This could have obscured any differences in the interpretation of different types of negation that might exist. The present experiment, therefore, uses a design in which participants answer only one question for each scenario. This should avoid the risk of participants perceiving similarity between questions and seeking to provide consistent answers.

The previous experiment allowed comparison of each type of affixal negation with “not” negation. However, each type of affixal negation was presented using a different set of scenarios. This could have limited any potential conclusions had any differences between the affixal negation conditions been observed. The current experiment will, therefore, seek to create scenarios that can be used in affirmative, “not” negated and three affixally negated conditions (“non-”, “un-” and “iN-”). There are a small number of stem words that are equally acceptable when negated using each of these three prefixes (Horn, 1989). However, while “iN-” negation is highly lexicalised and is only understood when applied to certain words, “non-” and “un-” are both productive and readily understood as part of novel compounds (Mazzon, 2004). However, “un-” is less frequently used as a prefix with stem words which form part of frequently used “iN-” compounds (Zimmer, 1964). To find suitable words that would be readily understood, a shortlist of stem words which form part of a compound with “iN-” were drawn up. Questions using affirmative and “not”, “un-” and “non-” negated versions of the stem were tested using a small pilot group. This group were asked whether they readily understood the question (even if they would not use the negated form themselves). All proposals were readily understood when “non-” negation was used. Some proposed stem words were rejected because the “un-” formation was confusing to the pilot group.

The final key change in this experiment will be the use of a much larger sample with participants completing the experiment online. The large sample is required to obtain the same amount of data as each participant will be answering fewer questions in a between-participants design compared to the previous experiment. It also helps to identify significant differences where effects are small as we expect here. Conducting the survey online makes it more practical for a large, diverse group to be recruited and to complete the experimental tasks.

#### 2.4.1.1. Online Recruitment of Participants

This experiment sought a much larger number of participants than the prior experiment. Therefore, an online survey was used to enable sufficient participants to be recruited and participate. Participants were recruited and paid through the ‘Amazon Mechanical Turk’ service ([www.mturk.com](http://www.mturk.com)). This service advertises short tasks to a pool of ‘workers’ who can choose to complete the task for the financial credit offered. Paolacci, Chandler and Ipeirotis (2010) reviewed the results of an experiment using the Mechanical Turk, comparing it to the results of the same experiment completed using a sample recruited within an academic setting and a sample recruited through internet discussion fora. They concluded that participants recruited through the Mechanical Turk were a reliable source of experimental data in judgement and decision making. Buhrmester, Kwang and Gosling (2011) also reviewed the use of the Mechanical Turk to recruit participants for psychology research. They found that participants were more diverse than typical internet samples and much more diverse than samples recruited within academic settings. They similarly concluded that the quality of data provided by Mechanical Turk participants met or exceeded standards required for published research.

#### 2.4.2. *Predictions*

As with the previous experiment, the participants were given limited information about the scenarios under consideration. It was expected that participants would respond consistently with superadditive judgement (Macchi et al., 1999). This prediction suggests that the average response to the affirmative and “not” negated condition questions combined in each scenario will be less than 100.

The previous experiment found that people made lower estimates of the likelihood of a proposition where implicit negation was used than where explicit “not” negation was used. However, the expected differences between “not” negation and affixal negation were not observed. The present experiment has been designed to be more sensitive to the anticipated small differences between the interpretation of “not” negation and the different types of affixal negation. Therefore, it was expected that each probability rating task using affixal negation would give rise to lower average answers than the equivalent question using “not” negation.

Within the three different types of affixal negation tested, “non-” is generally interpreted as the most contradictory (Mazzon, 2004). Therefore “non-” is expected to reveal the highest probability rating of the affixal negation conditions and the closest responses to “not” negation. Negation using “iN-” is generally interpreted as the most contrary of the three prefixes tested (Mazzon, 2004). It is therefore anticipated that probability rating tasks using “iN-” will receive the lowest probability ratings. Probability rating tasks using “un-” are expected to get responses between the other two forms of affixal negation tested.

### **2.4.3. Method**

#### **2.4.3.1. Design**

This experiment used the type of negation used in the question (five levels: affirmative – no negation; ‘not’; ‘non-’; ‘un-’; “iN-”) and the scenario (five levels) as factors. Each participant saw five tasks, one for each scenario and one for each polarity condition in a Latin Square Confounded experimental design (Kirk, 1995).

#### **2.4.3.2. Participants**

Participants were recruited through the ‘Amazon Mechanical Turk’ system which paid them US\$0.25 to complete the experimental survey. After potential duplicate responses (based on the respondent having the same IP address) were removed, the sample consisted of 464 participants aged between 16 and 67 (median age 26, 37% female, 73% with English as their first language)

#### **2.4.3.3. Materials**

Five scenarios were developed which provided some information about a hypothetical group of 100 people. The first part of the scenario set out the context. The scenario then had four bullet point which each described a characteristic that ‘some’ of the 100 would share. These statements were intended to create ambiguity about how many of the 100 people would meet the criteria which would be asked about. This would leave the response to the participants’ instinctive judgement. For example:

*100 people live in their parents' house.*

- *Some of those people pay a commercial rate of rent.*

- *Some of those people also use a parent's car.*
- *Some of those people have parents who are away for most of the year.*
- *Some of those people also go on holiday with their parents.*

Five questions were prepared for each scenario asking how many of 100 hypothetical people in the scenario the participant would expect to satisfy a proposition. One for these five questions used an affirmative proposition and the other four used different forms of negation (“not” negation and affixal negation using “non-”, “un-” and “iN-”). For example, the following questions were prepared for the scenario above:

*Out of the 100 people, how many would you expect to be dependent?*

*Out of the 100 people, how many would you not expect to be dependent?*

*Out of the 100 people, how many would you expect to be nondependent?*

*Out of the 100 people, how many would you expect to be undependent?*

*Out of the 100 people, how many would you expect to be independent?*

A web-based survey was prepared (using [www.surveygizmo.com](http://www.surveygizmo.com)) with the scenarios and the questions. The survey asked each participant a single question from each scenario (so five experimental questions in total). The survey presented the scenarios in random order. Across the scenarios, participants were asked one question for each polarity condition. The polarity condition assigned to each scenario was randomly determined for each participant.

Full details of the materials used in this experiment are provided in appendix 1 (see §A1.2)

#### 2.4.3.4. Procedure

Prospective participants were directed to the web survey which started with an information page and a consent statement which had to be accepted to allow them to proceed. They were then asked for some demographic information before being given the instructions for the experimental questions. The experimental questions then followed – one to a page – and each participant had to answer five. They had to enter a number between 0 and 100 in response to each question before being able to move on. The final pages included a code to enter into the Mechanical Turk system to claim



payment, thanked participants for their time and provided a contact e-mail address for any questions. The five questions that each participant answered represented one using each scenario and one in each polarity condition. Participants were randomly assigned to one of the 120 different combinations of the 25 prepared questions which were available under these criteria. The question order was then also randomised for each participant.

#### 2.4.4. Results

Table 2.3, below, sets out the mean responses for each polarity condition across all five scenarios. As anticipated, all three forms of affixal negation led to lower scores than “not” negation. The mean responses to the affirmative and “not” negated responses sum to less than 100. This is consistent with participants making, on average, superadditive judgements (where their estimate of the probability of all possible scenarios sums to less than 1). However, we cannot look at whether individual participants made superadditive or subadditive judgements as no individuals responded to both affirmative and “not” negated conditions for the same scenario.

*Table 2.3 – Summary of responses in chapter 2, experiment 2*

Polarity Condition	Affirmative	“not” negation	<i>non-</i> affixal negation	<i>un-</i> affixal negation	<i>iN-</i> affixal negation
Mean Response	48.2 (23.8)	45.1 (24.7)	44.2 (24.7)	45.0 (24.2)	41.5 (23.7)
Mean of Standardised responses	0.14 (0.87)	0.09 (0.89)	-0.00 (0.90)	-0.01 (0.88)	-0.14 (0.90)

Mean response (standard deviation) to 1 decimal place

Mean standardised response (standard deviation) to 2 decimal places

Each participant only saw one task for each scenario and one for each polarity condition. For this reason, additive scores – as used in the previous experiment – were not appropriate. Instead, participant effects were factored out by standardising responses. The means and standard deviations of each participant’s five responses were calculated (these included a response for each scenario and each polarity condition). Each response was then standardised by subtracting the participant’s mean response and dividing the result by their standard deviation.

A 5x5 between participants ANOVA was conducted on the standardised responses with scenario and question polarity as the factors. This found significant main effects for scenario ( $F(4,2295) = 15.777, p < .001$ ), polarity ( $F(4,2295) = 5.189, p < .001$ ) and a significant interaction between scenario and polarity ( $F(16,2295) = 6.556, p < .001$ ). The main effect of polarity in this test could be attributed to a substantial difference between the affirmative condition and all of the negated conditions which is not salient to our hypotheses. A 5x4 between participants ANOVA with scenario and question polarity as factors – with the affirmative condition removed – was therefore conducted. This still showed a main effect of polarity ( $F(3,1836) = 2.901, p = .034$ ) but the effect size was very small ( $\eta_p^2 = .004$ ). With the affirmative condition removed there was also still a main effect of scenario ( $F(4,1836) = 23.471, P < .001$ ) but no interaction between scenario and polarity ( $F(12,1836) = 1.554, p = .098$ ).

To understand the nature of this main effect of type of negation (after the affirmative condition was removed), comparisons were made using Tukey's HSD test. These comparisons showed a significant difference ( $p = .009$ ) between responses when "iN-" negation was used (mean 0.31, standard deviation 0.87) and responses when "not" negation was used (mean 0.07, standard deviation 0.86). No other significant differences were found using Tukey's HSD comparisons.

#### 2.4.4.1. Responses to Each Scenario

The results above showed a highly significant main effect of the scenario condition. While we have no hypothesis about the effects of different scenarios, the pattern of results might help to show how the use of different types of negation is context sensitive.

Comparisons using Tukey's HSD test were performed following the significant main effect of scenario found in the standardised responses after the affirmative condition was removed. These comparisons showed significant differences ( $p < .001$ ) between responses to the fifth scenario (using stem word 'offensive') and each other scenario (means and standard deviations set out in table 2.4). No other significant differences were found using Tukey's HSD comparisons.

Table 2.4 – Mean responses to each scenario in chapter 2, experiment 2

Scenario (stem word used)	A ( <i>dependent</i> )	B ( <i>active</i> )	C ( <i>religious</i> )	D ( <i>polite</i> )	E ( <i>offensive</i> )
Mean Response	42.9 (25.8)	42.6 (23.3)	42.0 (21.9)	44.1 (21.7)	52.5 (26.9)
Mean of Standardised responses	-0.18 (0.91)	-0.18 (0.82)	-0.20 (0.86)	-0.02 (0.85)	0.39 (0.91)

Mean response (standard deviation) to 1 decimal place

Mean standardised response (standard deviation) to 2 decimal places

The overall standardised results are in the anticipated pattern (responses to the affirmative condition are higher than all of the negated response; the negated responses decrease in the anticipated order: “not”; “non-”; “un-”; “iN-”). Scenarios A, B and C are all very close to this pattern individually. For Scenario A, responses when “un-” was used were higher than for all other types of negation. For scenario B, responses to “iN-” questions were higher on average than responses when “non-” or “un-” were used. For scenario C, responses when “un-” was used were higher than when “non-” was used. All of these differences which contradicted the anticipated pattern of results in the first three scenarios were very small and t-tests showed that they were not significant.

Standardised responses to scenario D were very close together. Although some of the mean responses did not fit the anticipated pattern (for example, responses when each form of affixal negation was used were higher than when *not* was used) t-tests were not significant for any of these differences.

As observed above, post-hoc tests showed that responses to scenario E were significantly different to each of the other scenarios. The pattern of results was quite different to that of the other scenarios. Responses to all of the negation conditions were uniquely higher than responses to the affirmative condition. Responses to questions where “non-” or “un-” were used were higher than responses where “not” negation was used.

#### **2.4.5. Discussion**

The observed main effect of polarity, which remained after the affirmative condition was removed, supports the basic hypothesis that the use of different types of negation in a proposition affects the perception of the probability of that proposition. The greatest difference between the negation conditions (and the only difference significant in post-hoc tests) was between “iN-” and “not” negation. This is consistent with the specific prediction that “iN-” would give rise to the lowest responses to probability rating tasks of the negation conditions and “not” the highest. This supports the expectation that using more contrary forms of negation will normally lead to lower estimates of the likelihood of a proposition than when more contradictory forms are used. The previous experiment showed that implicit negation based on contrary opposites yielded lower likelihood estimates than “not” negation. These results build on this finding by showing that similar differences are also observed when using different types of explicit negation.

However, there are several features of these results which do not directly support the hypothesis which need further consideration: The predicted differences between the other forms of affixal negation (“non-” and “un-”) and “not” negation; The inconsistency in the pattern of results when certain scenarios were used; and, the small size of the observed effects.

No significant differences were found between responses to tasks using “not” negation and “non-” and “un-” affixal negation (or between “non-” and “un-” negation). It was anticipated that response to both other types of affixal negation would be lower than “not” negation. Of the three forms of affixal negation tested, “non-” represented the most contradictory and least contrary form (Zimmer, 1964). We would, therefore, expect responses to be closest to those for “not” negation. In this context, it is unsurprising that no difference was found between “not” negation and affixal negation using “non-”.

Responses to questions using “un-” affixal negation were expected to fall between responses to questions using “non-” negation and responses to questions using “iN-” negation. It is possible that these effects were very small. The difference between responses to “not” negated questions and “un-” negated questions were expected to be

smaller than the difference between responses to “not” negated questions and “iN-” negated questions. The observed difference between responses to “iN-” negated and “not” negated questions was very small. It is therefore possible that any difference between responses to “not” and “un-” negated questions was too small to detect.

Another reason why questions using “un-” negation may not have elicited the expected differences, is that the scenarios may not have used this negator as participants would normally encounter it. The stem words used in the questions were chosen to work in affirmative, “not” negated and three affixally negated forms (“non-”, “un-” and “iN-”). As explained above, the “iN-” prefix is highly unproductive in modern English and it is only normally used with a narrow set of stem words in lexicalised compounds (Zimmer, 1964; Mazzon, 2004). Therefore, appropriate stem words for use in the scenarios for the current experiment were selected from words to which the “iN-” prefix is routinely applied. The other prefixes used, “non-” and “un-”, are both readily productive in modern English and people understand their use with a wide range of words (Zimmer, 1964; Mazzon, 2004). When piloting potential scenarios, the “non-” forms were readily accepted, but some other proposed scenarios were rejected because the “un-” forms confused people. Zimmer (1964) observed that the learned appropriateness of the “iN-” prefix to a given stem word might suppress the productivity of the “un-” prefix (even where the word otherwise had all the characteristics of a word that would be readily negated by “un-”). This may have made the words negated using “un-” seem unusual to participants – affecting their normal interpretation of this form of negation.

Looking at the pattern of results for each scenario may provide some support for the interpretation. Notably, responses to scenario B are close to those anticipated. However, responses to the question where ‘inactive’ was used were very slightly higher on average than responses when ‘nonactive’ or ‘unactive’ were used. This may be because, although ‘inactive’ is a widely used word, it is may normally be used to describe inanimate objects rather than people. In the context of describing people, participants may have seen ‘unactive’ as more natural, facilitating a response closer to expectations.

This experiment provided further evidence that context is important in how people interpret negations. In particular, scenario D (‘polite’) showed very little difference between the polarity conditions and scenario E (‘offensive’) reversed the normal pattern

of responses following affirmative and negated questions. A key difference between the first three scenarios ('dependent', 'active' and 'religious') and scenario E ('offensive') is that the former have positive or neutral emotional valence while the latter has negative valence. Jespersen (1917) argued that affixal negation was only used with stems that have positive emotional valence based on the perception that it is used to create negative expressions like 'unhappy' but not positive expressions like 'unsad'. Others have since shown that affixal negation can be used with negative stems although this may be less common (Zimmer, 1964; Horn, 1989; Mazzon, 2004). It may be this distinction that led to the different pattern of results observed with this scenario.

Only a small effect from different types of negation was found. It may be that such effects are always very small. However, the present experiment also only found a small difference between responses to affirmative and "not" negated conditions. A large effect would often be expected between these conditions if a person had reason to believe that the affirmative proposition was highly likely or highly unlikely (with the negation of it correspondingly unlikely or likely). As we did not see such an effect, it is possible that the small differences observed are related to the scenarios which were designed to create ambiguity about the likelihood of affirmative and negative conditions. Different scenarios might, therefore, elicit a larger difference in responses.

Finding differences in the probability ratings assigned to negated propositions depending on the type of negation used may affect how people reason with those propositions (e.g. Ohm & Thompson, 2006).

Therefore this experiment has built on our understanding of how different types of negation affect estimates of likelihood by showing how responses when extremely contrary "iN-" negation is used differ from when contradictory "not" negation is used. However, questions remain about whether a similar effect can be found when "un-" is used as a negator, how content effects interact with the use of affixal negation and whether effect sizes can be increased.

## **2.5. Experiment 3**

The previous two experiments suggested that people give different probability ratings to propositions based on the type of negation used in the proposition. However, these

experiments may not have been sensitive enough to detect potential differences between people's estimates of the probability of a proposition using "not" negation and the probability of the same proposition using "un-" affixal negation instead. This experiment seeks to provide a further test for this anticipated difference. Also, this experiment seeks to provide an initial test of how people understand conditional statements that use affixal and "not" negation. It looks at how people rate the plausibility of conditionals that use affirmative, "not" negated and "un-" negated propositions in their antecedent and consequent clauses.

The previous experiment showed that differences between the perception of "not" negation and affixal negation using "iN-" could be observed using probability rating tasks (Oakford, Chater & Grainger, 1999). However, any difference between "not" negation and "un-" negation may have been obscured by the scenarios used. The previous experiment's scenarios used stem words which readily negated with "iN-". As the availability of negation using "iN-" sometimes suppresses the productivity of "un-" as a negator (Zimmer, 1964; Mazzon, 2004), participants may have found the application of "un-" negation unusual. The present experiment will, therefore, use scenarios for which negation using "iN-" is normally inappropriate and with which the "un-" prefix is routinely used.

The previous two experiments have also provided scenarios which are designed to create ambiguity and which provide no overall sense of whether the test proposition is likely to be true or false for a given hypothetical person. This may have reduced observable differences between the negation conditions by clustering responses around 50%. In the current experiment, the scenarios' context statements will seek to imply a difference between the likelihood of affirmative and negated conditions. The stem words that are the subject of negation will also all have a positive or neutral emotional valance. As observed in the prior experiment, negating negative stems may affect their interpretation.

The experiments in this chapter aim to provide an initial test of the impact of different types of negation on estimates of a propositions likelihood. This is motivated by prior research which has shown that the likelihood of propositions used in conditional statements effects reasoning using those statement (e.g. Oakford, Chater & Larkin,

2000). Factors that affect the perceived probability of propositions are therefore expected to affect how people reason using conditionals using those propositions. It is therefore relevant to consider the effect on people's perception when propositions are presented as part of conditional rules rather than in isolation. The second part of this experiment will provide an initial test of how the use of different types of negation affects how people perceive conditional statements.

The perceived plausibility of a conditional statement is a feature which affects people's willingness to draw inferences using the statement (e.g. Quinn & Markovits, 2002). People might judge plausibility based on factors affecting the likelihood of the propositions in the conditional (De Neys, Shaeken & D'Ydewalle, 2002). This experiment will, therefore, look at how participants' perceptions about the plausibility of a conditional are affected when antecedent and consequent clauses are varied between affirmative, "not" negated and "un-" negated conditions.

### **2.5.1. Predictions**

As before, it is anticipated that probability ratings of affirmative propositions will be higher than ratings of propositions that use negation. It is also expected that probability ratings for propositions using "not" negation will be higher than for those using "un-" negation.

The second part of the experiment looks at how participants rate the plausibility of conditional statements which use antecedent and consequent clauses that are systematically varied between affirmative, "not" negated and "un-" negated versions. In line with the first part of the experiment, we might expect affirmative clauses to be perceived to be more likely than "not" negated clauses – which will, in turn, be perceived as more likely than "un-" negated versions. For a consequent to be true if the antecedent is true, the consequent must occur with at least the same frequency as the consequent. Conditional statements with high probability antecedents and low probability consequents are therefore likely to be interpreted as less plausible than statements with low probability antecedents and high probability consequents. The following pattern of results is therefore predicted:



- Conditionals with negated antecedents (i.e. lower probability) will have lower plausibility ratings when the consequents are affirmative (i.e. higher probability) rather than negated;
- Similarly, conditionals with “un-” negated antecedents will have lower plausibility ratings when consequents are affirmative or “not” negated; and
- Conditionals with “un-” negated consequents (lowest probability) will be rated less plausible than those with “not” negated consequents which, in turn, will be rated as less plausible than those with affirmative consequents (highest probability).

This pattern may be mediated by participants’ interpretations of the propositions relationships with each other, but no specific predictions are made.

### **2.5.2. Method**

#### **2.5.2.1. Design**

The first part of this experiment with probability rating tasks used polarity as the independent variable (three levels: affirmative; ‘not’; and, ‘un-’). All participants responded to a question with one polarity condition randomly selected for each of nine scenarios presented.

The second part of the experiment used the antecedent and consequent polarity of the presented rule the factors (each with three levels: affirmative; ‘not’; and, ‘un-’). All participants were presented with one of nine questions chosen randomly for each of four scenarios.

#### **2.5.2.2. Participants**

As in the previous experiment, participants were recruited through the ‘Amazon Mechanical Turk’ system to receive a small payment (in this case US\$0.13 each to complete a short survey). After responses from the same IP address were excluded (to avoid the risk that the same participant responded twice), 413 people aged between 16 and 68 participated (38% female, 49% with English as their first language). A small number of participants entered text in response to some of the questions requiring numerical responses and these answers have been excluded (none provided a clear numerical answer in text form). The total number of responses reported below is

therefore slightly lower than the number expected if all of the participants had submitted a complete set of responses.

#### 2.5.2.3. Materials

For the first part of the experiment, nine scenarios were generated that provided context for the probability rating questions. These stated the situation of 100 people and then made four further statements. The statements were designed to indicate that more people would meet the criteria of the following probability rating question in the affirmative condition than in the negative conditions. This was achieved by including two statements beginning ‘Many of those people...’ and ‘Several...’ consistent with the affirmative proposition, one statement beginning ‘Some...’ which was mildly inconsistent with the affirmative proposition and a statement beginning ‘A few...’ that was strongly inconsistent with the affirmative proposition. For example:

- 100 people were asked how they felt about their neighbours.*
- Many of those people often spoke with their neighbours.*
  - A few of those people actively disliked their neighbours.*
  - Several of those people sometimes socialised with their neighbours.*
  - Some of those people didn't often see their neighbours.*

These scenarios were followed by a probability rating question in one of three polarity conditions (one question was randomly assigned for each trial). For example, one of the following questions based on the stem-word ‘friendly’ was used each time the scenario above was given:

- *Out of the 100 people, how many would you expect to say their neighbours were friendly?*
- *Out of the 100 people, how many would you expect to say their neighbours were not friendly?*
- *Out of the 100 people, how many would you expect to say their neighbours were unfriendly?*

The second part of the experiment used eight of the scenarios used in the first part. Two scenarios from the first part were combined into each of four scenarios. For example:

*100 people were asked how they felt about their neighbours.*

- Many of those people often spoke with their neighbours.*
- A few of those people actively disliked their neighbours.*
- Several of those people sometimes socialised with their neighbours.*
- Some of those people didn't often see their neighbours.*

*The same 100 people are asked how they behave in a room full of new people.*

- Many of those people would try to make conversation.*
- A few of those people would find the situation so difficult they would want to leave the room.*
- Several of those people would want to use the opportunity to make new friends. - -*
- Some of those people would stay quiet.*

For each of the four scenarios, nine conditional rules were prepared and one of these rules was presented at random in each trial using the scenario following the statement that 'An observer looks at the 100 people and suggests the following rule'. These conditionals represented three antecedent polarity conditions and three consequent polarity conditions. For example, for the scenario above, the following conditionals were prepared:

- If a person is friendly, then the person is social*
- If a person is friendly, then the person is not social*
- If a person is friendly, then the person is unsocial*
- If a person is not friendly, then the person is social*
- If a person is not friendly, then the person is not social*
- If a person is not friendly, then the person is unsocial*
- If a person is unfriendly, then the person is social*
- If a person is unfriendly, then the person is not social*
- If a person is unfriendly, then the person is unsocial*

Full details of the materials used in this experiment are included in appendix 1 (see §A1.3).

#### 2.5.2.4. Procedure

Participants took part through a web-based survey. They confirmed their consent to participate and completed demographic questions before being given instructions for the probability rating tasks. Participants were presented with nine tasks, one for each probability rating scenario, one on each page in random order. For each scenario, the survey presented one of the three polarity conditions at random and the participants had to enter a number between 0 and 100 in response to the given question before they were allowed to move on. After the probability rating tasks, participants were given instructions for the plausibility rating questions. Participants were presented with each of the four plausibility rating scenarios on a page each in random order. For each plausibility rating scenario, they were given one of the nine possible conditional rules (three antecedent polarities combined with three antecedent polarities) at random. Participants were asked to indicate how plausible they thought the conditional rule was, given the context provided, using a Likert type scale with eight levels. After completing their tasks, they were thanked for their time and invited to return to the Mechanical Turk system and enter a code for payment.

### 2.5.3. Results

#### 2.5.3.1. Probability Rating Tasks – Overall Responses

Table 2.5 sets out the mean responses given to each question in the first part of the experiment. This shows that the overall mean responses and mean responses for five of the nine scenarios were in line with the hypotheses that the probability of affirmative propositions would be rated higher than the probability of the negative propositions and, between the negative propositions, the “not” negated propositions would have higher probability ratings than the “un-” negated propositions.

Table 2.5 – Responses to probability rating tasks in chapter 2, experiment 3

		Polarity		
		Affirmative	“not” negated	“un-” negated
Scenario’s stem word	overall*	57.6 (40.8)	37.4 (23.0)	36.0 (22.7)
	believing	61.1 (20.3)	37.7 (22.2)	38.8 (23.1)
	friendly*	57.4 (21.3)	35.9 (23.6)	35.1 (23.1)
	grateful*	67.6 (82.8)	38.3 (23.7)	32.1 (23.9)
	happy	59.9 (22.6)	38.0 (22.4)	38.4 (22.9)
	romantic	67.5 (61.8)	31.7 (22.0)	34.1 (22.7)
	stylish*	57.0 (23.5)	39.0 (25.2)	33.7 (21.7)
	wise*	50.5 (24.4)	38.4 (20.8)	36.2 (22.2)
	charitable	40.5 (22.8)	40.6 (24.7)	40.7 (22.0)
	social*	55.4 (22.2)	37.7 (21.5)	34.0 (21.3)

Mean response (standard deviation) to 1 decimal place

\* Mean responses to these scenarios (and the overall mean responses) were in line with the hypothesised pattern of results.

Participants’ responses were standardised by subtracting from each response the participant’s mean response and dividing by the standard deviation of the participant’s responses. Unless otherwise stated, subsequent results described here are based on these standardised scores rather than the raw responses. This standardisation leads to responses to the scenario based on the stem ‘happy’ also taking the hypothesised order, with the mean standardised affirmative score higher than the mean standardised ‘not’ score which is, in turn, higher than the mean standardised ‘un-’ score. Therefore, once the scores have been adjusted for participant biases in this way, six of the nine scenarios led to results in the hypothesised direction.

A 3x9 between participants ANOVA was performed with polarity and scenario as the factors. This found main effects for scenario ( $F(8,3650) = 3.029, p=.002$ ) and polarity ( $F(2,3650) = 316.223, p<.001$ ) and an interaction between scenario and polarity ( $F(16,3650) = 6.832, p<.001$ ). Our main focus was on whether there is a difference between the two negative polarity conditions, so a 2x9 between participants ANOVA was conducted using polarity (with only ‘not’ and ‘un-’ as levels) and scenario as factors. This found a main effect for scenario ( $F(8,2413) = 3.267, p=.001$ ). The main effect of polarity was not quite significant ( $F(1,2413) = 3.155, p=.076$ ). No significant interaction between scenario and polarity was found. As well as being non-significant, the polarity effect was small ( $\eta_p^2 = .001$ ).

#### 2.5.3.2. Probability Rating Tasks – Responses to Each Scenario

As we have seen above, six of the nine sets of mean standardised scores for each scenario exhibited difference in the predicted direction between each polarity condition. As we have previously observed, context is important to participant's ratings of likelihood and it is likely that each scenario will have influenced any negation effect observed.

For the three scenarios where the mean responses did not show the predicted pattern, two (based on stem words 'believing' and 'romantic') did show the anticipated difference between responses to the affirmative condition and each of the two negated conditions but not between the two negated conditions. However, the difference between responses to the negated conditions for each scenario was not significant ( $t(265) = -0.418, p = .676$ ;  $t(261) = -0.496, p = .620$ ). The final scenario where the mean responses did not display the predicted pattern (based on the stem word 'charitable') did not even show the expected difference between the affirmative condition and the negated conditions with almost identical mean responses given to questions in each polarity condition. Looking just at responses to questions on the six scenarios where the mean scores were in the anticipated order (based on stem words 'friendly', 'grateful', 'happy', 'stylish', 'wise' and 'social'), a 2x6 between participants ANOVA using polarity (with only 'not' and 'un-' as levels) and scenario as factors found a significant main effect for polarity ( $F(1,1608) = 7.874, p = .005$ ) but no main effect for scenario or interaction between polarity and scenario. The size of this main effect for polarity was larger than before the three scenarios described above were excluded ( $\eta_p^2 = .004$ ).

When each scenario was tested individually, only one set showed a significant difference between scores for the negated conditions (the scenario based on the stem 'grateful' ( $t(266) = 2.877, p = .002$ , one-tailed).

#### 2.5.3.3. Probability Rating Tasks – Possible Other Factors

The high proportion of participants who do not have English as a first language and the random allocation of polarity condition to each trial (without reference to the other trials that participants had responded to) may have been factors in these results.

More than half of the participants (51%) reported that English was not their first language. This reflects the global use of the Mechanical Turk system used to recruit participants (although it is much higher than for the previous experiment which also used the Mechanical Turk system). The difference between different types of negation in English was expected to create subtle differences in the perception of the questions and therefore responses. non-native English speakers might be less likely to pick up on and understand these differences in implication. A 3x9x2 between participants ANOVA was, therefore, conducted with whether English is the participant's first language as an additional factor alongside polarity and scenario. This found no significant effect of language ( $F(1,3623) = 0.038, p=.846$ ) and no significant interactions with scenario ( $F(8,3623) = 1.455, p=.168$ ), with polarity ( $F(2,3623) = 1.310, p=.270$ ) or with scenario and polarity ( $F(16,3623) = 0.409, p=.981$ ). When the participants that did not speak English as a first language were removed from the sample, the overall pattern of results remained the same (with higher standardised probability ratings given on average to affirmative propositions than negated conditions and higher ratings for “not” negated questions than “un-” negated questions). The 2x6 between participants ANOVA using polarity (with only ‘not’ and ‘un-’ as levels) and scenario (with the three inconsistent with the anticipated pattern removed) as factors was rerun with this reduced sample. As before it found a significant main effect for polarity ( $F(1,784) = 6.334, p=.006$ , one-tailed) but no main effect for scenario or interaction between polarity and scenario. The size of this main effect for polarity was slightly larger than when this test was performed with the full sample ( $\eta_p^2 = .007$ ).

Polarity conditions (either affirmative, “not” negated, or “un-” negated) were randomly allocated to each trial as participants answered one question for each scenario. This meant that each participant answered a different number of questions in each polarity condition. For example, a participant may have been randomly allocated to five questions in the affirmative condition, two in the “not” negated condition and two in the “un-” negated condition. A different participant may have been allocated to one affirmative trial, five “not” negated trials and three “un-” negated trials. The expectation was that any effects of this random distribution would be evened out over the whole sample. To check whether this may have led to any effects, the sub-group of 23 participants who each answered three questions in each polarity condition was identified. The overall pattern of results for this sub-group was the same as for the

overall group (with mean standardised probability ratings highest for responses to affirmative questions and lowest for “un-” negated questions with responses to “not” negated questions in the middle). Again the 2x6 between participants ANOVA using polarity (with only ‘not’ and ‘un-’ as levels) and scenario (with the three inconsistent with the anticipated pattern removed) as factors was repeated with this small sub-group. A main effect for polarity ( $F(1,76)=3.085$ ,  $p=.043$ , one-tailed) was found as before and no main effect for scenario or interaction between polarity and scenario were found. Again, the size of this main effect of polarity was larger than when this test was performed with the full sample ( $\eta_p^2 = .029$ ).

#### 2.5.3.4. Plausibility Ratings of Conditional Rules

Participants used an eight-point scale to evaluate the plausibility of the conditional rules which were provided (the response options were ‘Completely implausible’, ‘Very implausible’, ‘Implausible’, ‘A little implausible’, ‘A little plausible’, ‘Plausible’, ‘Very plausible’, ‘Completely plausible’ presented in that order). Responses were converted to scores between 1 and 8, with 1 representing ‘Completely implausible’ and 8 ‘Completely plausible’. Mean responses are provided in table 2.6.

*Table 2.6 – Summary of responses to plausibility tasks in chapter 2, experiment 3*

		Consequent Polarity			Overall
		Affirmative	“not” negated	“un-” negated	
Antecedent Polarity	Affirmative	5.7 (1.7)	3.8 (2.0)	3.7 (1.9)	4.4 (2.1)
	“not” negated	3.9 (2.0)	4.8 (2.1)	4.8 (1.9)	4.5 (2.0)
	“un-” negated	3.7 (2.0)	5.1 (1.7)	4.7 (1.8)	4.5 (1.9)
	Overall	4.4 (2.1)	4.6 (2.0)	4.4 (1.9)	4.4 (2.0)
Mean response (standard deviation) to 1 decimal place					

The overall pattern of results suggests that, where the polarity of antecedent and consequent terms aligned (i.e. both affirmative or both using either form of negation), plausibility ratings are generally above average. However, where the polarity of antecedent and consequent terms differed (i.e. one is affirmative and the other uses a form of negation), plausibility ratings are generally below average.

A3x3x4 between participants ANOVA was performed with antecedent polarity, consequent polarity, and scenario as the factors. This found a main effect of scenario ( $F(3,1604) = 3.679$ ,  $p=.012$ ) and significant interactions between antecedent polarity



and consequent polarity ( $F(4,1604) = 45.406, p < .001$ ) and between antecedent polarity, consequent polarity and scenario ( $F(12,1604) = 5.017, p < .001$ ). To understand the interaction between affirmative and consequent polarity a series of independent samples t-tests were used with Bonferroni adjusted significance levels ( $p = .006$ ). Where consequent polarity was affirmative, an affirmative antecedent clause led to significantly higher plausibility ratings than both a “not” negated ( $t(356) = 8.595, p < .001$ ) or “un-” negated ( $t(351) = 9.730, p < .001$ ) antecedent. Where consequent negation was “not” negated, an affirmative antecedent clause led to significantly lower plausibility ratings than both a “not” negated ( $t(364) = -4.744, p < .001$ ) or “un-” negated ( $t(332) = -6.296, p < .001$ ) antecedent. Finally, where consequent negation was *un* negated, an affirmative antecedent clause also led to significantly lower plausibility ratings than both a “not” negated ( $t(367) = -5.297, p < .001$ ) or “un-” negated ( $t(379) = -5.248, p < .001$ ) antecedent. No significant differences were found between the two negated antecedent conditions in any consequent negation condition.

#### **2.5.4. Discussion**

This experiment has shown some very small significant effects when we focus on the difference between the two negative polarity conditions only when certain materials are excluded (scenarios that showed differences in unanticipated directions). The removal of these materials could be justified on the basis that they are insufficiently sensitive to the factor of interest (polarity) because of the scale of other content effects. However, this justification is limited and no firm conclusions can be drawn without substantial caveats and a need for further research. These small differences may indicate that the use of “un-” negation in a proposition could lead to significantly different estimates of the probability of the proposition compared to where “not” negation is used. Both forms of negation lead to lower probability ratings in their propositions than the affirmative case but the key, predicted, difference is that propositions using “un-” would be rated with lower probability than those using “not”. This is in line with the expectation that “un-” negation would lead to a more contrary implication and “not” a more contradictory implication. This experiment was designed to highlight this effect which was obscured in the earlier experiments which sought to look at a wider range of negative operators.

Finding a potential difference between probability ratings for propositions when “not” and “un-” are used also addresses one of the issues with the previous experiment where this predicted relationship was not observed. The result, therefore, builds tentatively on the prior two experiments by providing evidence that may be consistent with the overall hypothesis that using different types of negation affects people’s perception of the probability of a proposition. This may suggest that the use of different types of negation will affect how people respond to reasoning problems. For example, Oaksford et al.’s (2000) model suggests that if the probability of the propositions in a conditional differ, a person’s willingness to draw an inference based on that conditional will change.

However, this finding is still limited. The size of the potential effect was still very small and the effect of scenario was also still very significant. However, it is likely that some issues with this experiment continued to obscure the full significance and size of the effect.

Focusing on the three scenarios that did not provide the predicted pattern of results, the results suggest that these were not sufficiently sensitive to pick up the anticipated difference in negated conditions. In particular, it is worth reviewing the scenario which did not provide even the expected difference between affirmative and negated conditions. It may be that the participants did not relate the statements in the scenario (which referred to whether people would think a suspect guilty) with the probability rating questions (which asked how many people would be charitable, not charitable or uncharitable). While this relationship had apparently been understood by a small number of pilot participants, they may not have been representative of the wider experimental sample. This is why it may be reasonable to exclude these scenarios from the analysis but – as mentioned above – such analysis should be treated very cautiously.

The results also considered whether the proportion of participants who did not speak English as a second language and whether the distribution of the polarity conditions may have had an effect. In both cases, it looks like these did not affect the pattern of results but that the effect size might have been bigger if these had not been factors. There may be scope for a future study to be developed which further refines the current paradigm and provides a more sensitive test of the effect of using different types of negation on responses to probability rating tasks.

Looking at the effect of different types of negation in conditional statements on perceptions of the statements' plausibility, the results showed that the use of negation has an effect. However, the results did not show any significant difference between responses where negation was used based on the type of negation. The results, therefore, supported the prediction that conditionals with negated antecedents would have lower plausibility ratings when the consequents are affirmative rather than negated. However, the other specific predictions were not supported. From that pattern of results, it appears that participants' interpretations of the relationship between the propositions were strong and so if one proposition was true that made the other more likely and vice versa. This would have led to the observed interaction with higher ratings of plausibility where both antecedent and consequent were affirmative and where both antecedent and consequent were negated. Lower plausibility ratings were observed when one clause was affirmative and the other negated. The next chapter will consider the role of plausibility in conditional reasoning further.

## **2.6. General Discussion**

People conduct deductive reasoning over propositions. The interpretation of propositions is, therefore, an important part of the reasoning process. Understanding what people perceive about propositions that use negations should, therefore, help us understand the process of reasoning and may provide insight into how and why biases occur.

Perceived probability is one of the features of propositions which has already been shown to affect reasoning using that proposition (e.g. Oaksford et al., 2000). The experiments in this chapter have looked at whether using different types of negation in a proposition affects people's expectations of the probability that the proposition is true. Each experiment found some evidence that this was the case, although the size of the effects observed was small.

In the first experiment, a clear difference between responses was seen when implicit negation was compared to explicit negation. In the second experiment, an overall difference between people's probability ratings when "not" and "iN-" negation were used showed that different forms of explicit negation could have an effect. In the third

experiment, this effect was also demonstrated between “not” negation and affixal negation using “un-” where the effect was anticipated to be smaller.

Implicit negation has been widely used in previous reasoning studies (e.g. Evans & Lynch, 1973). It has already been observed that when “not” negation is used in place of implicit negation some of the reasoning biases associated with negation are affected. For example, Cheng and Holyoak (1985) found that matching bias in Wason’s (1968) selection task was reduced when explicit negations were used in place of implicit negation. The finding that using implicit negation in place of explicit negation in a proposition affects its perceived likelihood may provide further insight into why these differences occur. They suggest that effects such as matching bias may not be down to people’s inability to reason with negations but down to probabilistic reasoning with low probability propositions that use implicit negation. This supports the probabilistic model for reasoning in the selection task proposed by Oaksford and Stenning (1992).

The further findings that different types of explicit negation (affixal negation using “iN-” and possibly “un-” and “not” negation) can affect likelihood estimates of propositions suggests that further consideration needs to be given to the application of negation in logical reasoning tasks. It often appears assumed that negation works as a simple contradictory operator with clear, logical meaning and implication. However, studies of negation in language have shown that the interpretation of negation can be complex and it can carry different implications (Horn, 1989). This suggests that any account of negation biases in reasoning tasks needs to consider how the different implications of negation may have been incorporated into the participants reasoning process. As we have seen, different types of negation affect the probability ratings of propositions using negation. The probabilistic model (e.g. Oaksford & Chater, 2001) which explicitly considers the probability of the propositions that make up the conditional statements that are reasoned over provides a mechanism for incorporating such effects. It is not clear how alternative models of reasoning based on mental logics (e.g. Braine & O’Brien, 1991) or mental models (e.g. Johnson-Laird, Byrne & Schaeken, 1992) can do this.

In the final experiment, the effect of different types of negation in conditional statements was explored. While it did not show any significant effects of different types

of negation, it did show how the use of negation has a significant effect on people's perception of the plausibility of a conditional. These differences may affect how people reason with a given conditional – and their willingness to draw inferences from it (e.g. Quinn & Markovits, 2002).

There is some evidence that the final experiment could be further refined to allow a more sensitive test of how different forms of negation affect the probability ratings of propositions. A further study would want to ensure that all scenarios provided a context that participants' found relevant to the probability rating question and evenly distribute questions with each polarity condition to each participant. It may also want to restrict the sample to prospective participants with English as a first language. However, for the purposes of the current study, the results in this chapter demonstrate that different types of negation can influence people's probability estimates. Previous research has shown that manipulating the probability of propositions within conditionals affects reasoning with those conditionals (e.g. Oaksford et al., 2000). The next chapter will, therefore, seek direct evidence of an effect of type of negation in conditional reasoning.

### **3. Different Types of Negation in Conditional Inference**

The previous chapter showed that different types of negation, like “iN-” and “not”, could have different effects on estimates of likelihood (albeit with small effect sizes). The probability assigned to propositions is a feature often associated with deductive reasoning over the propositions (e.g. Oaksford, Chater & Larkin, 2000). In this chapter, I will, therefore, look directly at whether the use of different types of negation leads to different responses when participants evaluate statements of conditional reasoning. I will use an extended version of Evans’ (1977) negation paradigm in conditional inference evaluation tasks.

#### **3.1. Evans’ (1977) Negation Paradigm**

As set out in §1.1, conditional inference evaluation tasks require participants to evaluate whether a given inference is true or false. The inference consists of a major premise (the conditional statement), a minor premise (whether the antecedent or consequent is true or false) and a conclusion (also whether the antecedent or consequent is true). As discussed in §1.2, when asked to evaluate a series of MP, DA, AC and MT inferences, all forms are typically endorsed more than half the time. Responses for MP inferences typically approach the ceiling, with MT yielding fewer endorsements but more than AC which, in turn, is more widely endorsed than DA (Schroyens, Schaeken, & d’Ydewalle, 2001). However, these results can be varied when materials are changed. For example, Evans and Handley (1999) found that when abstract materials are used, MP endorsement still approaches ceiling (95%) but MT endorsement (58%) is lower than AC endorsement (79%). Evans and Handley (1999) still reported that DA endorsement was lowest (32%).

A widely used evaluation task paradigm was proposed by Evans (1977) and has subsequently been described as the ‘negations paradigm’ (Oaksford and Stenning, 1992). Experiments using this paradigm present MP, DA, AC and MT inferences with each of four forms of conditional statement (providing a total of sixteen combinations of inference and conditional form). These four forms systematically varied the conditional statement’s antecedent and consequent terms between affirmative and negative polarities. These forms are described as AA, NA, AN and NN with the first letter setting

out whether the antecedent is affirmative or negative and the second letter setting out whether the consequent is affirmative or negative. The four forms are:

- $p \rightarrow q$  (AA)
- $\neg p \rightarrow q$  (NA)
- $p \rightarrow \neg q$  (AN)
- $\neg p \rightarrow \neg q$  (NN)

Using this paradigm Evans (1977) observed a negative conclusion bias (discussed in §1.2.1.2) when conditional statements of the form ‘if  $p$ , then  $q$ ’ were used (although not when the ‘ $p$  only if  $q$ ’ form was used). For DA, AC and MT inferences, endorsement levels were higher when the conclusion presented was negative (for MP, endorsement approached ceiling when the conclusion was both affirmative and negative).

The lack of negative conclusion bias observed when Evans (1977) used an alternative form of conditional statement highlights the importance of interpretation of the inference in the generation of negation effects. Showing that participants do interpret different conditional forms in different ways does suggest that if negations used in the conditionals are interpreted differently, there may be differences in the negation effects observed. Thompson and Mann (1995) suggested that “only if” statements are understood as having a more necessary relationship than “if, then” statements and a greater sense that one of the propositions must proceed the other.

Other experiments have sought to explore other factors that could influence the level of endorsement of different inferences. For example, Cummins, Lubart, Alksnis and Rist (1991) showed that people might endorse an inference in one context but reject it in a different context. They found that providing a context with a greater number of alternative causes (reasons why the consequent might occur that are not the current antecedent) or more disabling conditions (events which would prevent the truth of the consequent even where the antecedent occurred) reduced the level of inference endorsement.

Classical logic does not take account of context and when reduced to symbolic logic the temporal order of propositions described by Thompson and Mann (1995) does not

influence the truth function of a conditional statement. Therefore, whether the conditional is interpreted as a conditional or as a biconditional, classical logic does not provide an account for people endorsing different numbers of inferences depending on context.

Experiments using Evans' (1977) negations paradigm have been used to evaluate several models of reasoning. In particular, the mental models theory which implies that participants tend to represent affirmative propositions in internal models more easily (Johnson-Laird, Byrne & Schaeken, 1992). Evans (1993) suggested that, while this might provide a plausible explanation of responses to Wason's (1968) selection task (in which participants pick cards that match their internal models), it did not predict negative conclusion bias in evaluation tasks. Evans (1993) said that the mental models approach would predict an affirmative premise bias which had not been observed. Evans and Handley (1999) again found evidence of a negative conclusion bias (for DA and MT but not MP and AC) but no affirmative premise bias. Evans and Handley (1999) did find limited evidence for an affirmative premise bias when the materials used in evaluation tasks were very similar to those used in Wason's (1968) selection task suggesting that observed effects were potentially quite complicated. They suggested amendment to the mental models theory to better account for the data. Thompson and Mann (1995) used mental models theory to explain their analysis of the interpretation of "only if" conditional statements. They incorporated an indicator of temporarily prior events into the mental models approach and showed that their interpretation of "only if" would not lead to negative conclusion bias in the same way as statements using the "if, then" form.

Evans' (1977) negations paradigm has also been used to evaluate the probabilistic approach to explaining reasoning biases. Oaksford et al. (2000) developed a probabilistic account of conditional inference with reference to the contrast class account of negation (as previously discussed in §1.3.3). This model assumed the notion implied by the contrast class account of negation (discussed in §1.4.3) that negated propositions would normally have a higher probability than their affirmative contradictory (Oaksford & Stenning, 1992). Oaksford et al. (2000) showed the model was a good fit with data from a meta-analysis of previous conditional inference experiments and was supported by two experiments that manipulated the probabilities



of the propositions used directly. Oaksford and Chater (2007) looked at how well the probabilistic approach could account for results using Evans' (1977) negation paradigm when a different form of conditional rule was used. They suggested that different interpretations of the form of the rule and different ordering of propositions would affect the utility placed on different types of inference using decision theory. They showed that the results are consistent with facilitation of more useful inferences (those with higher utility) and suppression of less useful inferences. Oaksford et al. (2000) also look at how the probabilistic approach can account for negation effects. They created a probabilistic model which suggested that the likelihood a given inference would be endorsed is proportional to the conditional probability that the conclusion is true given the premises. This model was a good fit for their data.

### **3.1.1. Extending the Negation Paradigm**

The negations paradigm (Evans, 1977) has been used to test participants' evaluation of inferences with and without negations using different types of materials. This includes abstract (e.g. Evans, 1977) and more naturalistic materials (e.g. Oaksford et al., 2000). It also includes the use of both "not" negation and implicit negation (e.g. Evans, 1983). It has not previously been used to compare the effects of different types of explicit negation on inference endorsement levels. In order to compare the use of "not" negation and a different type of explicit negation, the range of conditional statement forms used needs to be expanded to systematically vary antecedent and consequent by three conditions: affirmative ('A'), "not" negated ('N<sub>1</sub>') and alternatively negated ('N<sub>2</sub>'). This produces nine forms of conditional sentence (the first letter represents the antecedent polarity and the second, the consequent polarity): AA, AN<sub>1</sub>, AN<sub>2</sub>, N<sub>1</sub>A, N<sub>1</sub>N<sub>1</sub>, N<sub>1</sub>N<sub>2</sub>, N<sub>2</sub>A, N<sub>2</sub>N<sub>1</sub> and N<sub>2</sub>N<sub>2</sub>.

The experiments in this chapter will compare "not" negation with affixal negation using "un-" as well as using affirmative terms in conditional inference tasks. Negation using *not* has been widely used in prior experiments using the negations paradigm (Evans, 1977). The experiments in chapter 2 demonstrated that different forms of negation can have different effects on people's probability estimates of propositions using negation. Negation using "un-" only showed a small effect when certain materials were used (a larger effect was seen with affixal negation using "iN-"). However, the prefix "un-" is more productive in modern English than "iN-" (and other negative affixes like *dis-* and

*-less*; Zimmer, 1964). It can be applied to a wider range of concepts and used to create novel compounds (Mazzon, 2004). It is, therefore, more relevant to everyday negation using modern English and easier to accommodate into experimental tasks without looking odd to participants or narrowing the range of materials available for the experiment.

As discussed in §2.2.1, explicit negation using the prefix “un-” creates a proposition with a more contrary interpretation than were the prefix “non-” or negation using “not” used. However, negation using “un-” is less contrary than negation which uses the praxis “iN-” (Zimmer, 1964; Horn, 1989). Previous models of reasoning based on the probabilistic approach have successfully modelled the data based on the assumption that negated propositions are high probability drawn from contrast class theory (Oaksford et al., 2000). This may be appropriate where highly contradictory “not” negation is used. However, more contrary interpretations of negation are likely to lead to a proposition with the negation being interpreted as lower probability. As “un-” negation typically takes a more contrary interpretation than “not” negation (but less than other forms of negation, like “iN-”) we might expect to see reduced negation effects when it is used (but still see some evidence of the previously observed biases).

### **3.2. Causal Relationships and Plausibility**

The experiments in this chapter will use different types of negation in conditional sentences. They will look at the effect these differences have on participants’ willingness to endorse inferences based on these conditional statements. Using different types of negation can alter people’s perception of a proposition’s likelihood as shown in chapter 2. It could affect people’s perception of the plausibility of a proposition or conditional statement using the proposition. It could also relate to people’s beliefs about the likelihood that the conditional describes a causal relationship. This chapter will, therefore, consider how people might look at the conditional as a causal relationship. It will also consider the impact of the people’s perceptions of the plausibility of propositions and conditional relationships between propositions on inference judgements.

### **3.2.1. The Conditional as Causal Relationship**

Some factors that could affect the perceived plausibility of conditional statements have been shown to affect the level at which people make or endorse inferences that use them. The plausibility of a conditional relationship is normally related to the perception that the relationship is causal, i.e. that ‘if  $p$ , then  $q$ ’ implies that  $q$  happens as a result of  $p$ .

As mentioned above, Cummins et al. (1991) showed that people were less likely to endorse an inference where they had been provided with a greater number of alternative causes (‘enablers’) or more disabling conditions (‘disablers’). In relation to a conditional ‘if  $p$  then  $q$ ’ (e.g. ‘if I turn the key, then the car starts’), enablers are alternatives to  $p$  that would also lead to effect  $q$  (e.g. hot-wiring the car) and disablers are things that might prevent  $p$  leading to effect  $q$  (e.g. no petrol). Cummins (1995) built on this in two experiments. Cummins’ (1995) first experiment used materials that had been selected based on pre-testing by a group that generated potential enablers and disablers in relation to the presented conditionals. The conditionals chosen for the experiment fell into four groups based on the pre-test: those with a high number of both enablers and disablers; those with a low number of both enablers and disablers; those with a high number of enablers and low number of disablers; and those with a low number of enablers and high number of disablers. Cummins’ (1995) experimental task again showed that people’s willingness to endorse inferences was affected by the number of enablers and disablers that could be generated for the conditional. She found that people’s evaluation of MP and MT inferences was more affected by the number of enablers (more enablers reduced certainty in the inferences), whereas DA and AC inferences were more affected by the number of disablers (more disablers reducing certainty in these inferences). This showed that people’s reasoning with a conditional was affected by factors that would affect the strength of the causal relationship described by the conditional. In a second experiment, Cummins’ (1995) showed that people were less certain in endorsing inferences using the conditional ‘If it rains, then the streets would be wet’ based on a scenario that would be familiar to people (and for which they would be easily able to generate alternative causes and disablers) than a conditional that used essentially the same form of scenario but in a foreign setting (‘If it

thardrons, then the streets will be sticky’ based on an alien planet which experiences ‘thardronning’ making the streets sticky in a manner similar to the ‘raining’ we experience making streets wet). This suggested that some conditional relationships – where people have no context to evaluate the likelihood of the propositions or the strength of the causal relationship – are interpreted logically. This also demonstrated that where people have information about the context, this affects their willingness to rely on the conditional when drawing inferences.

Ahn, Kalish, Medin and Gelman (1995) presented participants with potential causal relationships and investigated how people sought to understand and evaluate those relationships. They observed that people looked at the relationship presented and typically asked questions which tested hypotheses about the relationship or potential mechanisms through which the relationship might operate. Such results led Ahn and Kalish (2000) to argue that people perceived causal relationships to have a sense that one event necessarily leads to another. This distinguishes such relationships from correlations. They argued that where such a necessary relationship exists, people believe there a mechanism underpinning it. They argued that, although people are capable of identifying and testing relationships between events that regularly co-occur through induction, abduction is more likely to be relevant and useful to real-world tasks. Abduction is the development of the best available explanation for a relationship based on evaluation of alternative hypotheses. This proposal implies that a person’s perception of the plausibility of a relationship (which must be a part of the judgement of best available explanation for the relationship) is central to their willingness to draw inferences from causal relationships.

### ***3.2.2. Context and Plausibility in Causal Conditionals***

The use of naturalistic or thematic materials has been seen to reduce negation biases otherwise observed in reasoning tasks (e.g. Cheng & Holyoak, 1985). Essentially, when working with naturalistic materials, people appear to rely on existing world knowledge to inform the task. For example, Oaksford and Stenning (1992) argued that naturalistic material may be more amenable to everyday approaches to reason – which they suggest includes access to contrast classes for negated propositions.

As discussed above, knowledge of alternative antecedents to a causal conditional can reduce people's willingness to make inferences relying on the conditional. Chan and Chau (1994) provided supplemental conditionals of the form 'if  $r$  then  $q$ ' alongside conditional statements of the form 'if  $p$  then  $q$ ' and looked at the effect this had on participants drawing inferences using the latter conditional. Their results suggest that people are less willing to draw inferences of any type where the additional antecedent,  $r$ , is perceived to be more salient than the antecedent,  $p$ , to the consequent,  $q$ . Providing this additional information about the factors that the consequent is conditional on therefore appears to have changed people's perception of the relationship set out in the conditional over which they are asked to reason. This also shows that people use their prior knowledge (for example, of the salience of  $p$  and  $r$  to  $q$ ) where it is available in reasoning to alter their perception of the conditional statement. The use of prior knowledge to evaluate the rule may not be relevant where materials are abstract (materials such as those used by Evans, 1977, like 'If the letter is G, then the number is not 9').

Context to propositions can also be provided through information about how many times the events described occurred or didn't occur in different circumstances (contingency information). White (2000) provided participants with contingency information to explore how they would identify an event's cause. He found that people sought narrative explanations describing causal relationships (like conditionals) in preference to explanations based on simple co-variation between two events. The desired explanations were as complete as possible where information about all common factors and enabling factors was available (particularly when the level of confirmatory instances described by the contingency data was higher).

Quinn and Markovits (1998) compared people's perception of the strength of the relationship between antecedent and consequent in a conditional and their confidence in the inferences that they drew using it. Their participants were more certain of the AC and DA inferences that they made (suggesting a greater willingness to make them) where they perceived a strong association between the antecedent and consequent terms. In a further study, Quinn and Markovits (2002) provided participants with causal conditional relationships and context statements which altered the number of available alternative causes. These context statements changed the likelihood of a given

antecedent being responsible for a consequent event. Again they showed that AC and DA inferences were endorsed with more certainty where the plausibility of the conditional relationship was high (i.e. the number of available alternative causes was low).

To consider the effects of different types of negation in conditional reasoning, the current study must use somewhat naturalistic materials (while abstract tasks can use explicit and implicit negation it is hard to apply different forms of explicit negation and ensure they are universally understood). Given this, the current study will need to consider the context and plausibility of the conditionals used as part of the experiments.

### **3.2.3. Causal Models**

Causal Bayesian Networks (CBNs) are a type of Bayesian network (a probabilistic graphical model that sets out random variables and their conditional dependencies via a directed acyclic – i.e. not cyclic and therefore finite – graph). CBNs provide a way of illustrating the mental representations and processes that may underlie causal reasoning (Sloman, 2005). They treat causal dependences that people believe are operative as basic (Pearl, 2000) and are represented with nodes representing Bayesian random variables and arrows which run from cause to effect.

Fernbach and Erb (2013) provide recent CBN models (see figure 3.1) which show the relationship between  $p$  and  $q$  in a causal conditional ‘if  $p$  then  $q$ ’. In their models for MP and AC:  $W_p$  represents the strength of the causal relationship set out by that conditional;  $d_i^*$  represents the strength of potential disablers; and,  $W_a$  represents the strength of alternative causes that also result in proposition  $q$ . Fernbach and Erb (2013) argued that it was not simply the number of alternative causes which lessened  $W_p$  but also the strength of the alternatives. An increased number of alternative causes might reduce the strength of a causal relationship – and reduce the level of inferences people are willing to make based on the relationship. However, they proposed that a single strong alternative might have a greater effect than multiple weaker (or less likely) alternatives.

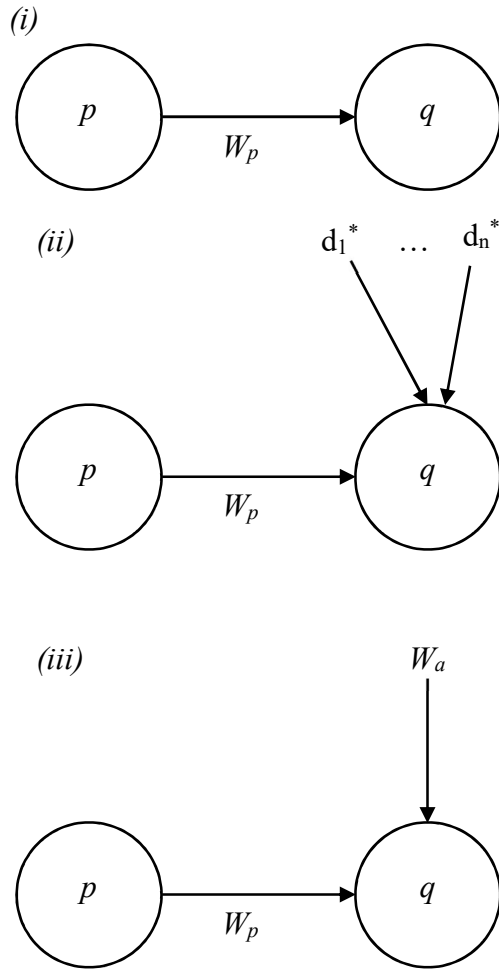


Figure 3.1 – Fernbach and Erb's (2013) CBNs for (i) simple MP, (ii) extended MP and (iii) AC inferences.

Fernbach and Erb's (2013) initial, 'simple', CBN model for MP inferences treats the relationship as straightforward, with a single parameter: the strength of the causal relationship,  $W_p$  (as set out below with  $a$  representing potential alternative causes of  $q$  – the disablers to the proposed causal statement).

$$\text{MP} = P(q \mid p, \neg a) = W_p$$

Fernbach and Erb's (2013) model for AC was parameterised using the base probability,  $P_p$ , and causal power,  $W_p$ , of the antecedent clause and the strength of alternative causes for the consequent clause,  $W_a$ :

$$\text{AC} = P(p \mid q) = 1 - (1 - P_p) \frac{W_a}{P_p W_p + W_a - P_p W_p W_a}$$

Fernbach and Erb (2013) noted that the data from Cummins' (1995) first experiment (discussed above, §3.2.3) showed some characteristics which were not entirely predicted by the quantity of enablers or disablers. For example, AC was more endorsed where there were many enablers and few disablers than where there were few enablers and few disablers although the quantity of enablers and disablers was similar in both conditions. Fernbach and Erb (2013) used the conditionals from Cummins' (1995) first experiment, but asked participants to provide judgements of the prior probabilities of the causes in the conditions, the causal power of the cause and the strength of alternative causes. They used a sampling procedure to take these parameters and generate model predictions for AC and MP inferences using their AC and simple MP models. They found that their model predictions, which considered the strength of enablers, were highly correlated with the responses observed by Cummins' (1995) for AC. For MP, their initial model was less conclusive.

Fernbach and Erb (2013) conducted a further experiment which sought judgements about the likelihood and strength of disablers related to the conditionals used. They used an 'extended' MP model. This model considered the role of disablers in reducing the strength of the causal relationship described by the conditional (the potential alternatives to the causal conditional being true).  $P_{di}$  represents the likelihood and  $d_i^*$  represents the strength of the  $i$ -th disabler,  $d_i$ :

$$MP = P(q \mid p, \neg a) = W_p = 1 - \left( \sum_{i=1}^n P_{di} d_i^* - \sum_{i,j:i < j} P_{di} d_i^* P_{dj} d_j^* + \sum_{i,j,k:i < j < k} P_{di} d_i^* P_{dj} d_j^* P_{dk} d_k^* - \dots + (-1)^{n-1} \prod_{i=1}^n P_{di} d_i^* \right)$$

Using this model, Fernbach and Erb (2013) did demonstrate a good fit for the MP judgements made by their participants in their second experiment with a larger and more significant correlation than models based on the number of disablers alone and mental models theory.

In a final experiment, Fernbach and Erb (2013) looked at causal and "non-causal" conditionals (like 'If it is over 80 degrees then a marathon runner will sweat' and 'If the marathon is on a Tuesday then a marathon runner will sweat' respectively). Given the high likelihood of the consequent, a conditional probability model that predicts the level of inferences based on 'if  $p$  then  $q$ ' will be related to  $P(q \mid p)$  predicts that inferences



using both conditionals will be similarly endorsed. However, Fernbach and Erb (2013) showed that casual conditionals led to more highly endorsed inferences than non-causal conditionals. This supports their model which looks at conditional inference as a causal relationship.

The current study will consider what models like that proposed by Fernbach and Erb (2013) might suggest when different types of negation are used in reasoning problems.

Ali, Chater and Oaksford (2011) looked at how CBNs could provide an algorithmic account of conditional reasoning that was consistent with a probabilistic computational account. They compared the predictions based on a CBN account with those made by the mental models account (e.g. Johnson-Laird, Byrne & Schaeken, 1992), the main competitor in providing an algorithmic level account of conditional reasoning. They considered certain predictions that probabilistic CBN models and mental models made about people's estimates of the likelihood of propositions used in conditional inferences. For each task, they presented two conditionals with different antecedents but the same consequent. Participants were either told that the consequent was the case (the 'consequent' condition) or not told anything further about the consequent (the 'non-consequent' condition). They were asked to rate the likelihood of one of the antecedents presented both before and after being told that the other antecedent is true and measured the difference between these ratings. They also varied the type of conditionals used. Some tasks using conditionals which placed cause before effect (e.g. 'if I have no oil, then my car breaks down', called the 'causal' cases). Others using conditionals which placed effect before cause (e.g. 'If it is warm outside, then it is sunny', the 'diagnostic' cases).

Ali, Chater and Oaksford (2011) made a series of predictions based on the mental models and CBN accounts. Using a CBN representation then providing two diagnostic conditionals and the fact that the consequent is true, knowing that one of the antecedents is true does not affect the likelihood of the other (which is also a consequence of the consequent) leading to the prediction of no differences between likelihood ratings. However, when the truth of the consequent has not been established, the knowledge that one consequence of the antecedent happens will increase the likelihood that the consequent is true and therefore a person would anticipate an increased likelihood rating

between the questions before and after the alternative antecedent is asserted to be true. Where a causal conditional is used, then, if the truth of the consequent is not known, the likelihood of a given cause is likely to be similarly low both before and after finding out that an alternative cause is true. However, knowing that the consequent is true will increase the likelihood that a given cause is true only for this likelihood to be suppressed when an alternative cause is established leading to a substantial negative difference in the likelihood ratings. Consideration of mental models led to a different set of predictions. They considered initial and full mental models. Initial mental models do not change with confirmation that an alternative antecedent is true in any of the experimental conditions so suggest no difference between likelihood ratings in each task should be anticipated. Similarly, where the truth of the consequent is known, then the full mental models generated before and after the truth of the alternative antecedent are established are the same, so no differences in the perceived likelihood of the subject antecedent are anticipated (for both causal and diagnostic conditional). Where the truth of the consequent is not known, then the full mental model representation does change when the truth of an alternative antecedent is asserted – the number of possibilities is reduced and the number in which the subject antecedent is true increase as a proportion of the number of possibilities. This would lead to an anticipated positive difference between likelihood responses in each task. This effect will be greater for causal conditional than diagnostic conditionals (because a greater proportion of possibilities where the target antecedent are true remain in the causal case than the diagnostic case after the alternative antecedent has been asserted and cases inconsistent with this have been removed from the full mental models).

Ali, Chater and Oaksford (2011) conducted two experiments to test whether the CBN or mental models approaches better accounted for responses. The first treated the presence of the statement that the consequent was true as a between-subjects variable and asked for two likelihood ratings for each task as described above. The second treated the presence of the statement that the consequent was true as a within-subjects variable and only asked for a single view of how likelihood had changed after the alternative antecedent was asserted in each task. In both experiments, a larger number of CBN predictions were shown to be true than those derived from either initial or full mental models' theory.

### **3.3. Experiment 1**

This initial experiment using conditional inference tasks is intended to explore whether the extended negation paradigm will show different levels of inference when different types of negation are used. It will also collect information about the plausibility ratings participants assign to the conditional statements used. This will help to consider any interaction between negation and plausibility effects on participants' willingness to endorse inferences using the conditionals. It will also help to identify scenarios where plausibility differences between negation conditions are minimised for use in a future experiment which seeks to focus only on negation effects. The results of this experiment were previously reported in Vance (2011, unpublished masters dissertation), but the discussion is updated.

As an initial test, this experiment will not use the full extended negation paradigm for its conditional inference evaluation tasks. Instead, only the antecedent term of the conditional statements will be varied systematically through the three polarity questions. This will allow a large number of different naturalistic scenario forms to be tested (twelve) with four different inferences (MP, DA, AC, MT) in a within-participants design with a manageable number of trials. By only varying the polarity of the antecedent term, each participant will only need to respond to 144 trials – rather than the 432 required if the full extended negation paradigm is used. The antecedent term was selected over the consequent term as the target of the polarity condition (with three levels: affirmative, “not” negated and “un-” negated) based on prior research which has shown a higher level of negation effects with the antecedent is varied than the consequent (e.g. Evans, 1977; Oaksford et al., 2000).

The second part of the experiment gathers plausibility rating information for all of the conditional statements required by the extended negations paradigm using the twelve scenarios (108 conditionals). This will provide information on the perceived plausibility of the 36 conditional statements used in the first part of the experiment. This will allow the results to this part of the experiment to be considered in the light of the plausibility ratings assigned to the conditional. By using the full extended negation paradigm, responses to the plausibility rating questions can be analysed to see if they show any systematic effects of using different types of negation in the antecedent and consequent

terms. Such an effect may suggest a potential mechanism by which negation in conditionals affects inferential reasoning responses. Collecting information about conditionals covering the full extended negation paradigm for all twelve scenarios will also help select a smaller set of scenarios for use in a further experiment using conditional inference evaluation tasks. It will allow the scenarios that show the least variability in perceived plausibility across polarity conditions to be selected. This will allow the next experiment to focus on the effects that different types of negation may have with lower risk that plausibility factors confound the result.

Each participant's plausibility ratings will be measured using a visual analogue scale. Participants will be presented with a line with each end labelled with opposite extreme conditions ('Extremely implausible' and 'Extremely plausible'). They will be invited to indicate where on the line they would site their perception of the plausibility of the conditional sentence given. Grant, Aithison, Henderson, Christie, Zare, McMurray, and Dargie (1999) suggest that such scales can provide increased sensitivity and replicability to the measurement of subjective perceptions compared to Likert type scales.

### **3.3.1. Predictions**

The use of contrary "un-" negation in the antecedent term is expected to create an interpretation of that term which behaves more like the affirmative case (despite having opposite meaning) than the case using highly contradictory "not" negation. Typically, the use of negation in antecedent terms is thought to provide a higher probability proposition than the affirmative alternative. Oaksford and Stenning (1992) and studies which reflect its contrast class account of negation (e.g. Oaksford et al., 2000) argue this is because the negation triggers generation of a contrast class of options that typically combines to be more probable than the single option provided in the alternative. The potential access to alternative options is also relevant to typical CBN models (e.g. Fernbach & Erb, 2013) where the strength of the conditional is affected by the perceived strength of alternative causal conditionals with different causes causing the same effect. While a "not" negated affirmative proposition may give rise to consideration of a contrast class – or alternatives to the affirmative proposition, the same affirmative proposition with "un-" negation which has a narrower implication. Therefore, while "not" negation is likely to lead to the same – high probability –

interpretation of the negation proposition seen in prior studies, “un-” negation will lead to a lower probability interpretation, more like the affirmative version of the proposition. For example, if someone is “not happy”, there are a wider range of alternative emotional states that might be considered in scope of the proposition (e.g. neutral, bored or stressed) than would be considered if the person was described as “unhappy”. While “unhappy” may still generate a range of alternatives larger than the affirmative proposition (e.g. “sad”, “angry”) these are also fully included in the wider scope of “not happy”. Therefore, even where an “un-” negated proposition covers a wider range of possibilities than the unnegated proposition, it will always have fewer than the “not” negated equivalent statement.

Therefore, we anticipate a pattern of inference endorsement where “un- ” negation is used on a proposition that is more like that where the affirmative proposition is used than the pattern seen where “not” negation is used on the proposition (although it might not completely match if “un-” negated statements still generate some alternatives). Any such effect is likely to be most observable where all the conclusions for a given inference have the same polarity (i.e. in this experiment for MP where they are all affirmative and DA where they are all negative). Such patterns may not be observable where conclusions have different polarities where might also expect to see the replication of the well-established negative conclusion bias (for example, Evans, Clibbens & Rood, 1995) when responses to inferences using “not” negation in their conclusions are compared to inferences with affirmative conclusions. This means that a higher level of endorsement of DA, AC and MT inferences is anticipated where the conclusion is “not” negated compared to when the conclusion is affirmative. As in previous studies, we do not anticipate seeing this negative conclusion bias in trials using MP inferences where all responses are likely to approach the possible ceiling obscuring any observable differences (as in, for example, Oaksford et al., 2000).

In this experiment, AC inferences using affirmative, “not” negated and “un-” negated conclusions will be used. As discussed above, we would expect those conclusions using “un-” to behave more like affirmative propositions than equivalents using “not” negation. We would, therefore, expect to see less negative conclusion bias demonstrated in AC trials which used “un-” in their conclusion than the equivalent trials using “not”.

Johnson-Laird and Byrne (2002) discussed the mental models approach to explaining negation effects. Where negations are used in a condition, these are incorporated into the relevant mental models. Where a proposition is then negated, they can lead to the need to negate a negation in the mental model to generate an affirmative solution. Johnson-Laird and Byrne (2002) argue that this requirement to process a double negation makes some affirmative conclusions harder to process and endorse than negative conclusions. Whether “not” or “un-” negation is used in the proposition, the requirement mental model would still require consideration of double negations (in some cases the “not un-” double negation). Therefore, the mental models approach would not anticipate any differences in inference responses between different types of negation.

Any effects will be moderated by the strength of the causal relationship perceived by the conditionals with different antecedent terms. For example, where there is a strong perceived causal link described by the relationship “if p then q” then the causal link described by “if not p then q” and “if un-p then q” is likely to be weak. This is likely to be reflected in high plausibility ratings for “if p then q” and low plausibility ratings for the similar conditionals with negated antecedents. A lower level of inference endorsement would be expected where the plausibility of the conditional is lower.

For the second part of this experiment, there is no specific prediction for the impact that different polarity conditions will have on the plausibility ratings given to the conditionals. This experiment will seek to find scenarios that minimise the differences in the plausibility of conditional statements when different polarity conditions are applied to the antecedent and consequent. Where differences in the plausibility ratings given to the conditionals used in the first part of the experiment, inferences using more plausible conditionals would be expected to have higher endorsement levels than inferences using less plausible conditionals. This is in line with prior findings that people are more willing to endorse conditional inferences where the conditional is more plausible (e.g. Quinn & Markovits, 1998, 2002).

### **3.3.2. Method**

#### **3.3.2.1. Design**

This experiment used a 4x3 within-participants design for its first part. The independent variables were inference type (with four levels: MP, DA, AC and MT) and antecedent polarity (with three levels: affirmative, “not” negation and “un-” negation). The dependent variable was the responses to each conditional inference provided, whether they endorsed the inference or not.

The second part of the experiment looked at the impact of different types of negation on plausibility using a 3x3x12 within-participants design. The independent variable factors were antecedent polarity and consequent polarity (each with three levels: affirmative, “not” negation and “un-” negation) and scenario (with twelve levels: one for each scenario). A visual analogue scale was used to capture participants’ responses to the question of how plausible they thought the given conditional relationship was. These were recorded as a number between 1 and 100 and used as a dependent variable.

#### **3.3.2.2. Participants**

A total of nineteen participants were recruited for this study. This represents an opportunity sample and all the participants were known to the experimenter. Participants were provided with an information sheet before being asked to confirm informed consent and were made aware that they could withdraw from the study at any time. The median age of participants was 35 (ages ranged from 24 to 57) and thirteen were female. English was the first language of seventeen participants – the other two were fluent in English. Participants were each tested individually using a laptop computer running E-Prime.

#### **3.3.2.3. Materials**

Twelve scenarios were prepared which could be used to form conditional statements and each form of inference in the different polarity conditions required. Each scenario consisted of a person’s name (e.g. “Andrew”), a first proposition (e.g. “predictable”, a statement which placed the proposition geographically and/or temporarily (e.g. “in the kitchen tomorrow”), a pronoun (e.g. “he”) and a second proposition (e.g. “clean”). The two propositions in each scenario were all single words which could be used

affirmatively or negated using both “not” and “un-”. The statement placing the proposition geographically and/or temporarily was intended to make the conditional seem less abstract without altering the logical implication. For example, the following DA inference was used with “not” negated antecedent:

*If Melanie is not predictable in the kitchen tomorrow then she will be tidy.*  
*Melanie is predictable in the kitchen tomorrow.*  
*Therefore Melanie won't be tidy.*

For the first part of the experiment, these scenarios were used to provide 144 inference statements which could be presented to participants. This represented each of the twelve scenarios being used in each of the four inference types (MP, DA, AC and MT) for each of three antecedent polarity conditions (affirmative, “not” negated and “un-” negated). The following forms were used to generate the inference statements for each type of inference:

- MP
  - *If [person's name] is [n1][first proposition] [statement of when and/or where] then [pronoun] will be [second proposition].*
  - *[person's name] is [n1][first proposition] [statement of when and/or where].*
  - *Therefore [person's name] will be [second proposition].*
- DA
  - *If [person's name] is [n1][first proposition] [statement of when and/or where] then [pronoun] will be [second proposition].*
  - *[person's name] is[n2] [first proposition] [statement of when and/or where].*
  - *Therefore [person's name] won't be [second proposition].*
- AC
  - *If [person's name] is [n1][first proposition] [statement of when and/or where] then [pronoun] will be [second proposition].*
  - *[person's name] will be [second proposition].*
  - *Therefore [person's name] is [n1][first proposition] [statement of when and/or where].*
- MT
  - *If [person's name] is [n1][first proposition] [statement of when and/or where] then [pronoun] will be [second proposition].*
  - *[person's name] won't be [second proposition].*
  - *Therefore [person's name] is[n2] [first proposition] [statement of when and/or where].*



The antecedent polarity was varied by altering the text in the [n1] and [n2] placeholders. For the affirmative condition, [n1] was blank and [n2] was “n’t”. The “not” negated condition used “not ” and the “un-” negated condition used “un” for [n1]. Both negation conditions left [n2] blank.

The second part of the experiment used conditional statements generated using the same twelve scenarios as the first part of the experiment. The basic form of the conditional statements was:

- *If [person’s name] is [n1][first proposition] [statement of when and/or where] then [pronoun] will [n2][second proposition].*

Nine different versions of this conditional form were generated for each of the twelve scenarios creating 108 conditionals to be tested. These nine versions represented all combinations of antecedent polarity in three conditions (affirmative, “not” negated and “un-” negated) and consequent polarity in the same three conditions. The [n1] placeholder was varied to create the antecedent negation conditions. It was blank for the affirmative condition, “not” in the “not” negated condition and “un” in the “un-” negated condition. The [n2] placeholder was varied to create the consequent negation conditions. It was “be ” for the affirmative condition, “not be ” in the “not” negated condition and “be un” in the “un-” negated condition.

Further information about the materials is included in appendix 2 (see §A2.1).

#### 3.3.2.4. Procedure

After participants agreed to participate, they were asked to sign a consent form and complete a short demographic questionnaire. They were seated in front of the computer monitor with a keyboard and mouse in front of them. The experiment began after they confirmed that they were comfortable and ready.

For the first part of the experiment, they were shown a screen providing instructions on responding to the conditional inference tasks. This explained that on each of the following screens, they would be shown an inference and they should press ‘A’ if they believed that the conclusion given followed from the relationship and circumstance provided as the premises and ‘L’ if they did not. The 144 conditional inference trials

were then presented one at a time in random order. Participants had to respond before moving onto the next trial.

Once the conditional inference tasks were completed, participants were provided with a screen setting out the instructions for the plausibility questions. On each screen that followed, they would be shown a conditional relationship and a line labelled 'Entirely implausible' on the left end, 'Neither plausible nor implausible' in the middle and 'Entirely plausible' on the right end. They were asked to move a cursor along the line with the mouse and click on the line to indicate how plausible they thought the statement given was. Once they had understood the instructions and were happy to proceed the plausibility tasks followed on 108 screens in random order. Participants had to click on the line on each screen to move onto the next task.

Once finished, participants were thanked for their time and given a chance to ask the experimenter any questions that they had about the study.

### **3.3.3. Results**

In the first part of the experiment, 79.3% of all inferences were endorsed overall. This was made up of 95.9% of MP, 73.8% of DA, 76.9% of AC and 70.7% of MT inferences. In the second part of the experiment, the mean plausibility ratings assigned to any given conditional statement ranged from 25.5 to 83.5 with a mean rating across all statements of 53.9.

#### **3.3.3.1. Conditional Inference – Levels of Endorsement**

Figure 3.2 shows the proportion of inferences of each type and in each antecedent negation that participants' endorsed. Two 4x3 within participants ANOVAs were conducted with inference type and antecedent negation as factors. The first looked at the proportion of trials that each participant endorsed inferences with each type of inference and antecedent negation. This found a significant main effect of inference type ( $F(3,54) = 6.321, p=.001$ ) and a main effect of type of antecedent polarity which approached significance ( $F(2,36) = 2.897, p=.068$ ). The second ANOVA compared the proportion of trials that were endorsed for each of the twelve scenarios with each type of inference and antecedent polarity. This found significant main effects of inference type ( $F(3,33) = 121.936, p<.001$ ) and antecedent polarity ( $F(2,22) = 4.112, p=.030$ ). This also found an

interaction between inference and antecedent negation which approached significance ( $F(6,66) = 2.116, p=.063$ ).

The specific prediction that, for the AC inferences, endorsement would be higher where the antecedent was “not” negated than where it was “un-” negated was tested using a paired samples t-test. This only approached significance ( $t(18) = 1.353, p=.097$ , one-tailed). A further series of twelve t-tests were conducted to investigate whether some scenarios might be more sensitive to the anticipated effect than others. This found significant a significant effect for scenario 7 ( $t(18) = 2.191, p=.042$ , one-tailed) and an effect approaching significance for scenario 9 ( $t(18) = 1.372, p=.094$ , one-tailed).

In order to look at whether changes to antecedent polarity had other effects on each type of inference, a series of one-way ANOVAs were conducted for each type of inference with antecedent polarity as the factor. The results of these ANOVAs approached significance for MP ( $F(2,36) = 2.939, p=.066$ ) and DA ( $F(2,36) = 2.636, p=.085$ ) inferences. No significant effect was found overall for MT or AC inferences.

A series of three 4x2 within participants ANOVAs were also conducted with inference (MP, DA, AC, MT) and antecedent polarity (each used two out of three of affirmative, “not” negated and “un-” negated) as factors. Inference had a main effect in each case ( $F(3,54) = 6.542, p=.001$ ;  $F(3,54) = 5.650, p=.002$ ;  $F(3,54) = 6.362, p=.001$ ).

Antecedent polarity had a main effect which approach significance when the affirmative and “not” negated conditions were compared ( $F(1,18) = 4.230, p=.055$ ) but not when the affirmative and “un-” negated conditions or the “not” negated and “un-” negated conditions were compared. There were no interactions between inference and antecedent polarity in any of these tests. For each inference type, a set of three planned comparison t-tests compared each antecedent negation condition with the others. None of these comparisons for any inference reached the required level of significance (Bonferroni adjusted significance level of  $p=.017$  was used).

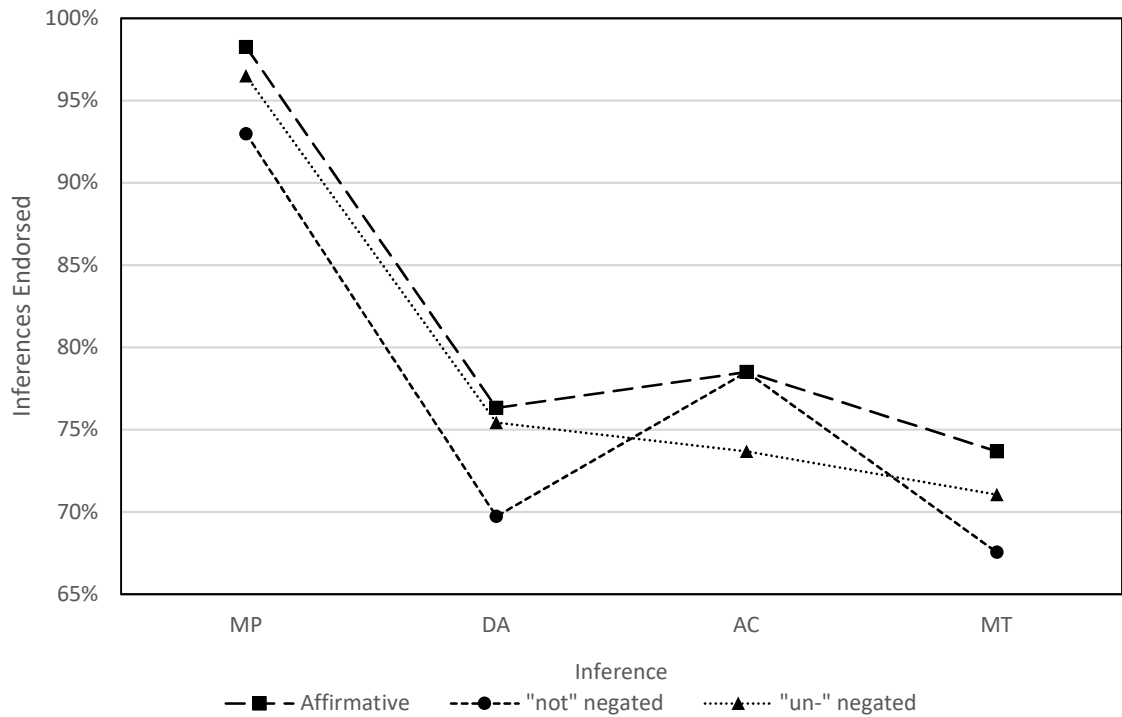
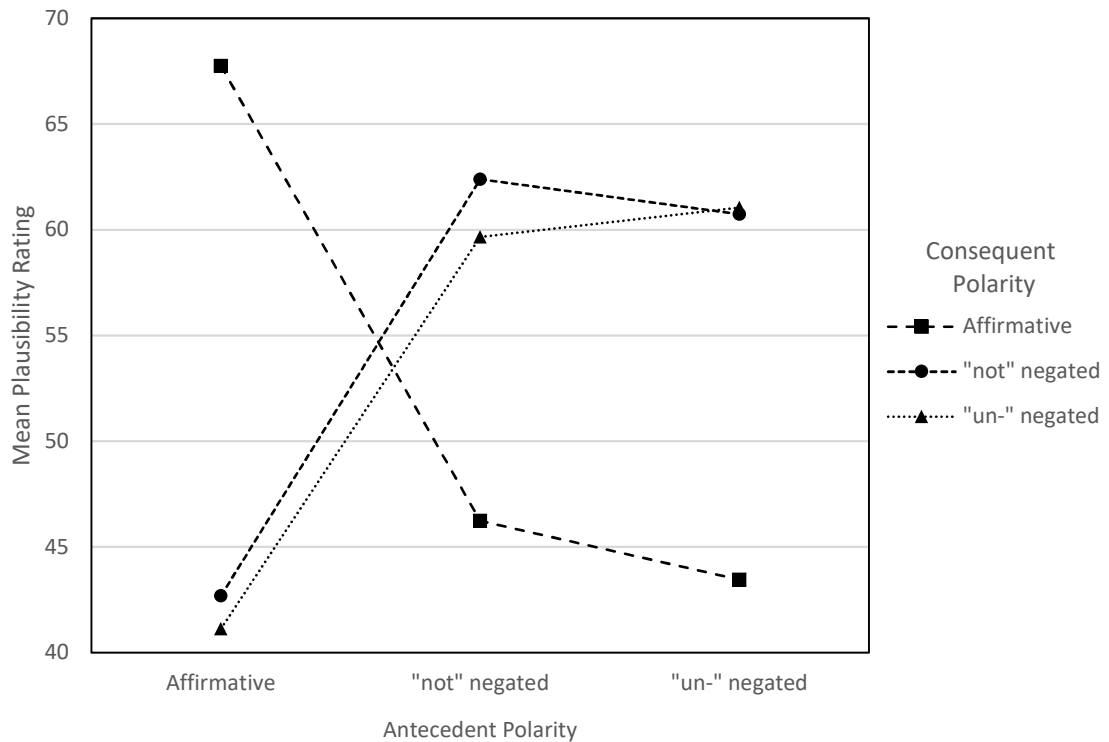


Figure 3.2 – Chart showing percentage of inferences by inference and antecedent polarity in experiment 1.

### 3.3.3.2. Effect of Polarity on Plausibility Ratings

Figure 3.3 shows the mean plausibility ratings assigned to conditional statements by antecedent and consequent polarity conditions. To look at any effects created by the variation in polarity of antecedent and consequent terms on the plausibility ratings assigned to the condition statements, a 3x3x12 within participants ANOVA was conducted. This ANOVA had antecedent polarity and consequent polarity as factors (each with three levels: affirmative, “not” negated and “un-” negated) and as well as scenario used (which had twelve levels). This found a significant interaction between all three factors ( $F(44,792) = 8.924, p < .001$ ). This also found significant interactions between consequent polarity and scenario ( $F(22,396) = 4.504, p < .001$ ), between antecedent polarity and scenario ( $F(22,396) = 2.050, p = .004$ ) and between antecedent polarity and consequent polarity ( $F(4,72) = 46.984, p < .001$ ). Main effects of antecedent polarity ( $F(2,36) = 8.765, p = .001$ ) and scenario ( $F(11,198) = 2.114, p = .021$ ) were also found.



*Figure 3.3* – Chart showing mean plausibility ratings by antecedent and consequent polarities in experiment 1.

To help interpret the interaction effect between the antecedent polarity and consequent polarity factors, each participants' responses across all 12 scenarios to each combination of the antecedent and consequent polarity conditions was combined to provide the participant's mean ratings for each polarity combination. Three post hoc comparisons were then carried out for each of the three antecedent polarity conditions, comparing the affirmative and "not" negated consequents, the affirmative and "un-" negated consequents and the "not" negated and "un-" negated consequents. Similarly, three post hoc comparisons were then carried out for each of the three consequent polarity conditions, comparing the affirmative and "not" negated antecedents, the affirmative and "un-" negated antecedents and the "not" negated and "un-" negated antecedents. These used a Bonferroni adjusted significance level of  $p=.002$ . Significant differences were found between affirmative consequents and each of the negated consequents for all three antecedent conditions and between affirmative antecedents and each of the negated antecedents for all three consequent conditions. No significant differences were found between the two negated consequent conditions for any antecedent condition or between the two negated antecedent condition for any consequent condition. The significant results are shown in table 3.1.

*Table 3.1 – Comparisons between antecedent and consequent polarity conditions*

Antecedent polarity	Test between polarities of consequents		
	Affirmative and “not” negated	Affirmative and “un-” negated	“not” negated and “un-” negated
Affirmative	t(18) = 8.017, p<.001	t(18) = 8.341, p<.001	NS
“not” negated	t(18) = -5.550, p<.001	t(18) = -4.610, p<.001	NS
“un-” negated	t(18) = -5.305, p<.001	t(18) = -5.483, p<.001	NS

Consequent polarity	Test between polarities of antecedents		
	Affirmative and “not” negated	Affirmative and “not” negated	Affirmative and “not” negated
Affirmative	t(18) = 7.033, p<.001	t(18) = 6.771, p<.001	NS
“not” negated	t(18) = -6.768, p<.001	t(18) = -5.543, p<.001	NS
“un-” negated	t(18) = -6.549, p<.001	t(18) = -6.316, p<.001	NS

NS = not significant; Bonferroni adjusted significance level of p=.002 used.

To consider how different scenarios might affect the impact that changing the antecedent and consequent polarity conditions will have on plausibility ratings, a 3x3 repeated measures ANOVA was conducted with antecedent polarity and consequent polarity as factors for each scenario. The results of these ANOVAs are set out in table 3.2.

*Table 3.2 – Effects of antecedent and consequent negation on plausibility of conditional*

		F-statistics for 3x3 ANOVAs		
		Antecedent Negation	Consequent Negation	Antecedent Negation x Consequent Negation
Scenario	Scenario 1	0.764	3.879**	11.431****
	Scenario 2	4.237**	10.617****	11.568****
	Scenario 3	4.888**	2.002	14.657****
	Scenario 4	16.350****	5.896***	33.256****
	Scenario 5	0.026	1.763	7.219****
	Scenario 6	0.023	2.875*	0.712
	Scenario 7	2.376	0.074	12.846****
	Scenario 8	3.274**	1.633	24.621****
	Scenario 9	6.807***	5.177**	32.510****
	Scenario 10	1.626	2.792*	9.943****
	Scenario 11	0.905	8.244****	14.047****
	Scenario 12	0.543	5.744***	0.503

\* p≤.1, \*\* p≤.05, \*\*\* p≤.01, \*\*\*\* p≤.001

### 3.3.3.3. Interaction Between Plausibility and Conditional Inference

Of the 108 conditional statements that were rated for plausibility in the second part of this experiment, 36 had been used in the first part of the experiment (those with affirmative consequent terms). To understand how the plausibility attributed to these statements may have influenced responses to the conditional inference tasks, the three conditionals with the highest mean plausibility ratings for each of the three antecedent polarity conditions were identified. These were scenarios 6, 11 and 12 with affirmative antecedents, scenarios 1, 4 and 8 with “not” negated antecedents and scenarios 4, 8 and 9 with “un-” negated antecedents. Responses to these tasks were combined into a group labelled ‘high plausibility’. Overall, 82% of responses in the high plausibility group endorsed the given conditional inference. Similarly, the three conditionals with the lowest mean plausibility ratings for each of the three antecedent polarity conditions were identified. These were scenarios 4, 8 and 9 with affirmative antecedents and scenarios 6, 11 and 12 for both “not” negated and “un-” negated antecedents. Responses to these tasks were combined into a group labelled ‘low plausibility’. Overall 78% of responses in the low plausibility group endorsed the given conditional inference.

To compare responses in the high and low plausibility groups, a 4x3x2 repeated measures ANOVA was completed with inference (MP, DA, AC and MT), antecedent polarity (affirmative, “not” negated and “un-” negated) and plausibility (high plausibility and low plausibility) as factors. This found significant main effects of inference ( $F(3,54) = 6.456, p=.001$ ) and plausibility ( $F(1,18) = 5.046, p=.037$ ). The mean number of inferences endorsed was higher for the high plausibility group than for the low plausibility group for every inference and antecedent polarity condition except those with MT inferences that had affirmative antecedent terms in the conditional.

The main aim of this experiment is to look at whether the use of different types of negation in conditional inferences affects inference rates. To consider whether there is a negation effect separate from any plausibility effects, separate analysis was conducted on the group of responses that used scenarios where there is no main effect of the antecedent polarity condition on plausibility ratings. As shown in table 3.2 scenarios 1, 5, 6, 7, 10, 11 and 12 showed no significant main effect of antecedent negation on plausibility ratings. Using only responses to these scenarios, a 4x3 repeated measures ANOVA with inference type and type of antecedent polarity as factors was completed on the number of inferences endorsed. This found a significant main effect of inference type ( $F(3,54)=6.248, p=.001$ ) but not antecedent polarity. For the AC inferences using only these scenarios, 46.5% of inferences with “not” negated antecedent terms (and conclusions) were endorsed compared to the lower level of 43.0% of inferences that used “un-” negated antecedents (and conclusions). However, this difference was not significant.

### **3.3.4. Discussion**

Overall levels of inference endorsement were high compared to previous studies using conditional inference tasks with Evans’ (1977) negations paradigm (for example, compared to the average results found in Oaksford et al.’s, 2000, meta-analysis of previous conditional inference experiments). This overall high level of endorsement appears to arise from higher than usual levels of DA, AC and MT endorsement. MP endorsement approaches ceiling at a similar level to that found in prior studies. A potential explanation for the high levels of endorsement is that the language used in the instructions was open and informal, describing the conditionals as ‘relationships’ rather



than ‘rules’ for example, and did not make reference to logic. This may have led more people than usual to endorse inferences when in doubt. Another factor was the variable proportion of each inference type that was presented with a negative conclusion.

The overall level of inferences in different antecedent polarity conditions for MP, DA and MT placed the level of inference endorsement where “un-” negation was used between the levels where affirmative propositions and “not” negation are used. While this difference was not observed when looking at the different levels of inferences for each inference, the ANOVAs conducted suggest that the key effect of antecedent polarity is a difference between affirmative and “not” negated conditions (and not a difference between affirmative and “un-” conditions or the two negation conditions). This is in line with the prediction that “un-” propositions would behave more like affirmative propositions in influencing inference levels. This pattern of results for MP and DA would not reflect any conclusion bias effects because all conclusions used in each of these inference tasks will have had the same polarity (all affirmative for MP and all “not” negated for DA). For MT inference, conclusions were “not” negated where the antecedent was affirmative and conclusions were affirmative where the antecedents had either negation condition. The pattern of results for MT (with higher endorsement of inferences with “not” negated conclusions than of inferences with affirmative conclusion) was therefore in line with the expected negative conclusion bias although this difference was not significant.

The predicted pattern, described above, did not hold for AC inferences where the level of endorsement of inferences with “un-” negated antecedents in the conditional (73.7%) was lower than where the antecedent was both affirmative and “not” negated (both conditions had an overall endorsement range of 78.5%). However, this may reflect the anticipated negative conclusion bias increasing endorsements of “not” negated conclusions (which coincide with “not” negated antecedents). This effect may have overridden the normal pattern we might expect to see when negative conclusion bias is not a factor. As anticipated, more AC inferences were made in the “not” negated condition than the “un-” negated condition. However, this difference was only significant for one of the twelve scenarios and only approached significance overall. A similar difference was observed when only the scenarios with less difference in plausibility ratings were considered but this also failed to reach significance. The

direction of difference observed, and that the difference approached significance overall, supports the hypothesis that negative conclusion bias would be greater when “not” negation was used in conclusions than when “un-” negation was used. However, the lack of significant result may indicate that any effect is very small and an experiment with a larger sample size or more sensitive materials is required to detect it. Another curious feature of the pattern of responses to the AC inferences is that inferences with affirmative conclusions were endorsed at the same level as those that used conclusions with “not” negated. This may be because of material effects.

The results of this experiment are therefore indicative that antecedent polarity may be a factor affecting the level of inferences endorsed in its own right in the manner predicted (and therefore support a probabilistic interpretation, e.g. Oaksford et al., 2000).

However, the failure to reach significance on most of the key anticipated differences means this conclusion must be treated with caution. The experiment has also not clearly demonstrated whether different types of negation can lead to different levels of negative conclusion bias (where the key results were also non-significant). What it has done is highlighted the role of the plausibility of a conditional in people’s willingness to endorse inferences which use it. When high plausibility conditionals were used in inferences, they were significantly more likely to be endorsed than when equivalent low plausibility conditionals were used. When only scenarios with no plausibility differences over the antecedent polarity condition were considered, no significant effects of antecedent polarity on the level of inference endorsements were found. The importance of the plausibility of the conditional is in line with prior research on the plausibility of conditional statements (e.g. Quinn & Markovits, 1998, 2002; Ahn et al., 1995). Given the scale of the plausibility effect, it may be difficult to detect any independent negation effects unless differences between the plausibility of scenarios is minimised.

Looking at the effect on plausibility ratings of changing antecedent and consequent term polarities in conditionals, there are some significant interactions with the scenario used. This suggests that the different scenarios, with different subject matter, have their meaning changed in different ways by the inclusion of negation. This is not surprising given the nature of the naturalistic scenarios used but suggests that further reflection is needed when scenarios are selected for further studies.

### 3.4. Experiment 2

This experiment fully implements the extended negations paradigm based on Evans' (1977) negations paradigm for conditional inference evaluation tasks. The experiment is intended to look at whether the different negation conditions had different effects on reasoning. The extended paradigm, therefore, incorporates an additional, alternative, form of negation, as a condition to the antecedent polarity and consequent polarity factors. This means when the polarity conditions are systematically varied, there are nine forms of conditional to test rather than four (as in Evans, 1977). Testing each conditional statement in all nine polarity conditions with four forms of inference (MP, DA, AC and MT) therefore requires 36 trials. To keep the length of the experimental survey manageable for participants, only two different scenarios are used in the experimental trials (generating 72 experimental tasks). As with the previous experiment, affixal negation using "un-" was the alternative form of negation used in antecedent and consequent terms along with affirmative versions of the terms and "not" negated versions.

This experiment also aims to look at whether negation effects independent of plausibility effects can be identified. This appears to require the use of conditional statements which have minimal variance in perceived plausibility over all polarity conditions. The previous experiment looked at the plausibility rating participants gave to the twelve scenarios used across the nine different forms required by the extended negations paradigm. Two scenarios showed no significant interaction effect on plausibility ratings by antecedent and consequent polarities (scenarios 6 and 12). Neither of these scenarios demonstrated a main effect of antecedent negation on plausibility ratings. They did show some main effect of consequent negation on the conditional's plausibility ratings, but this is much smaller than the interaction effects seen on the other ten scenarios. This experiment will, therefore, use these two scenarios with the aim of minimising the influence of conditional plausibility on the number of endorsements in different polarity conditions.

Only using two different scenarios in the experimental trials creates an issue that participants are likely to become very aware of terms of the scenario and potentially bored. They may pay less attention over time. Such order effects should not impact on

the results if the trials are randomised. However, to minimise the repetition and potential for boredom, this experiment also uses a set of non-experimental distraction tasks. These follow the same format as the experimental tasks and use scenarios from the prior experiment which were not used in the experimental tasks. These will double the number of inference task trials the participants have to complete to 144 with experimental and distraction tasks all presented in random order together.

The previous experiment had a fairly low number of participants. This will have limited the power of the experiment to identify small effects. This experiment will, therefore, take the form of an online study and seek to recruit substantially more participants to complete it.

### **3.4.1. Predictions**

This experiment intends to provide a further test of the predictions related to the potential effects of different types of negation that experiment 1 considered (see §3.3.1). In this experiment, both antecedent and consequent negation will be varied so a more complex pattern as the polarity of those terms interacts should be observed. As with the prior experiment, the use of “un-” negation in the antecedent term is expected to create an interpretation of that term which behaves more like the affirmative case than the case using “not” negation. Therefore, an overall pattern of results is expected that places the level of inference where “un-” negated propositions are used between the levels of inference where affirmative and “not” negated propositions are used. As before this expectation is in line with expectations from probabilistic (e.g. Oaksford et al., 2000) and CBN (e.g. Fernbach & Erb, 2013) accounts, but not mental models accounts (e.g. Johnson-Laird & Byrne, 2002).

As before, this overall pattern may not be observed where we would expect to see negative conclusion bias (e.g. Evans et al., 1995) effects. In this experiment, DA and MT inferences will have conclusions that are affirmative in some tasks and “not” negated in others. For these inferences higher endorsement is anticipated where “not” negation is used than where conclusions are affirmative. All three polarity conditions will be used in conclusions of MP and AC tasks. As in the previous experiment, it is predicted that there will be a lower level of endorsement of inferences using “un-” in the conclusion than the equivalent inferences using “not” in the conclusion. Inferences with

both forms of negative conclusion are expected to be endorsed more often than their equivalents with affirmative conclusions.

As before, it is likely that ceiling effects obscure any effects on MP inferences. This experiment is intended to be more sensitive to the potentially small effects predicted than the prior experiment. As this experiment seeks to control for the perceived plausibility of conditionals in different antecedent and consequent polarity conditions, the previously predicted (and observed) plausibility effects should be minimised.

### **3.4.2. Method**

#### **3.4.2.1. Design**

The experiment used a 4x3x3 within-participants design. The factors were inference type (MP, DA, AC and MT), antecedent polarity (affirmative, “not” negated and “un-” negated) and consequent polarity (affirmative, “not” negated and “un-” negated). The responses to each conditional inference task (whether the inference was accepted or not) were recorded as the dependent variable.

#### **3.4.2.2. Participants**

A total of 96 participants (51 female) took part in this study. The participants were aged between 15 and 82 (with a median age 35.5) and 79 had English as a first language. All completed the experiment online through [www.surveygizmo.com](http://www.surveygizmo.com).

The participants were recruited in two ways. The first group (37 participants) was an opportunity sample of people known to the experimenter and asked to participate and those recruited through adverts and links on a number of websites and forums. These participants were motivated by an offer to donate money to charities. Participants were asked which of three charities they wished to support and a total £200 donation was divided among the charities in the proportion that participants had selected each of them.

The second group (59 participants) were recruited using the Amazon Mechanical Turk service (see §2.4.1.1). Participants recruited in this manner were paid \$3 each for their participation.

There were 26 complete responses to the survey that were excluded from the results and from the counts provided above because we could not be confident that the participants were unique. The excluded responses were those that used IP addresses which matched other responses (indicating that the participants are in the same location – although that location may be a university of business with many unique users) and did not provide a unique e-mail address or demographic information. Where two or more responses from the same IP address were found with no distinguishing information, all responses from that IP address were excluded.

#### 3.4.2.3. Materials

Conditional inference tasks were prepared based on the two scenarios found in experiment 1 in which the rules had the fewest significant differences in plausibility ratings between the different polarity conditions (those referred to as scenarios 6 and 12 in the previous experiment). These scenarios will be referred to as scenarios 1 and 2 respectively in relation to this experiment and its analysis. As before, each scenario consisted of a person's name, first proposition, statement of when and/or where, pronoun and second proposition. The propositions were words that could be used affirmatively, negated using 'not' and negated using 'un'. For example, the following was used as an MT statement with affirmative antecedent and "un-" negated consequent:

*If Brenda is informed about the issue being discussed today then she will be unconcerned.*

*Brenda will be concerned.*

*Therefore Brenda isn't informed about the issue being discussed today.*

Each scenario was made into 36 tasks: nine tasks for each of the four conditional inference types with the antecedent and consequent term polarities each varied three ways. This generated a total of 72 experimental tasks. A further 72 distraction tasks which were not part of the experiment were also generated based on several other scenarios used in experiment 1. These tasks were intended to create more variety in the scenarios presented to the participants and reduce the frequency that they were asked questions about the same two scenarios. All 144 tasks were displayed in a random order to each participant. The tasks provided three statements setting out a relationship,

second premise and conclusion. The statements took the following forms for each inference type.

- MP
  - *If* [person's name] *is* [n1] [first proposition] [statement of when and/or where] *then* [pronoun] *will* [n2] [second proposition].
  - [person's name] *is* [n1] [first proposition] [statement of when and/or where].
  - *Therefore* [person's name] *will* [n2] [second proposition].
- DA
  - *If* [person's name] *is* [n1] [first proposition] [statement of when and/or where] *then* [pronoun] *will* [n2] [second proposition].
  - [person's name] *is* [n3] [first proposition] [statement of when and/or where].
  - *Therefore* [person's name] [n4] *be* [second proposition].
- AC
  - *If* [person's name] *is* [n1] [first proposition] [statement of when and/or where] *then* [pronoun] *will* [n2] [second proposition].
  - [person's name] *will* [n2] [second proposition].
  - *Therefore* [person's name] *is* [n1] [first proposition] [statement of when and/or where].
- MT
  - *If* [person's name] *is* [n1] [first proposition] [statement of when and/or where] *then* [pronoun] *will* [n2] [second proposition].
  - [person's name] [n4] *be* [second proposition].
  - *Therefore* [person's name] *is* [n3] [first proposition] [statement of when and/or where].

The polarity of the antecedent and consequent negation were varied for each inference by replacing the placeholder items (labelled 'n1', 'n2', 'n3' and 'n4') as shown in table 3.3.

Table 3.3 – Varying polarity conditions

Antecedent Polarity	Consequent Polarity	Item			
		n1	n2	n3	n4
Affirmative	Affirmative		<i>be_</i>	<i>n't</i>	<i>won't</i>
Affirmative	"not" negated		<i>not be_</i>	<i>n't</i>	<i>Will</i>
Affirmative	"un-" negated		<i>be un-</i>	<i>n't</i>	<i>Will</i>
"not" negated	Affirmative	<i>not_</i>	<i>be_</i>		<i>won't</i>
"not" negated	"not" negated	<i>not_</i>	<i>not be_</i>		<i>Will</i>
"not" negated	"un-" negated	<i>not_</i>	<i>be un-</i>		<i>Will</i>
"un-" negated	Affirmative	<i>un-</i>	<i>be_</i>		<i>won't</i>
"un-" negated	"not" negated	<i>un-</i>	<i>not be_</i>		<i>Will</i>
"un-" negated	"un-" negated	<i>un-</i>	<i>be un-</i>		<i>Will</i>

After each set of three statements, each task asked the participant to indicate whether they thought the conclusion followed from the premises or not by clicking on the relevant radio button and then clicking a 'Next' button to move onto the next question. The order of the two radio buttons used to indicate whether the participant endorsed the inference or not was randomised for each trial.

Further details about the materials used in this experiment are included in appendix 2 (see §A2.2).

#### 3.4.2.4. Procedure

Participants taking part in the study were directed to an initial webpage that provided an introduction to the study and ethics information. Participants were asked to click on a 'Next' button to indicate that they consented to participate. The next page provided full instructions and was followed by three practise tasks in the same format as the experimental and distraction tasks. When the practise tasks were completed, participants were presented with the instructions again and asked to click 'Next' to continue. The 144 experimental and distraction tasks were then presented in random order, one to a page. After they had completed the reasoning tasks, participants were asked to complete a demographic questionnaire. They were given the option of providing their e-mail address if they wanted to participate in future studies, asked to indicate which of three charities they wanted to support, given an opportunity to provide comments and feedback on the experiment and finally thanked for their participation.



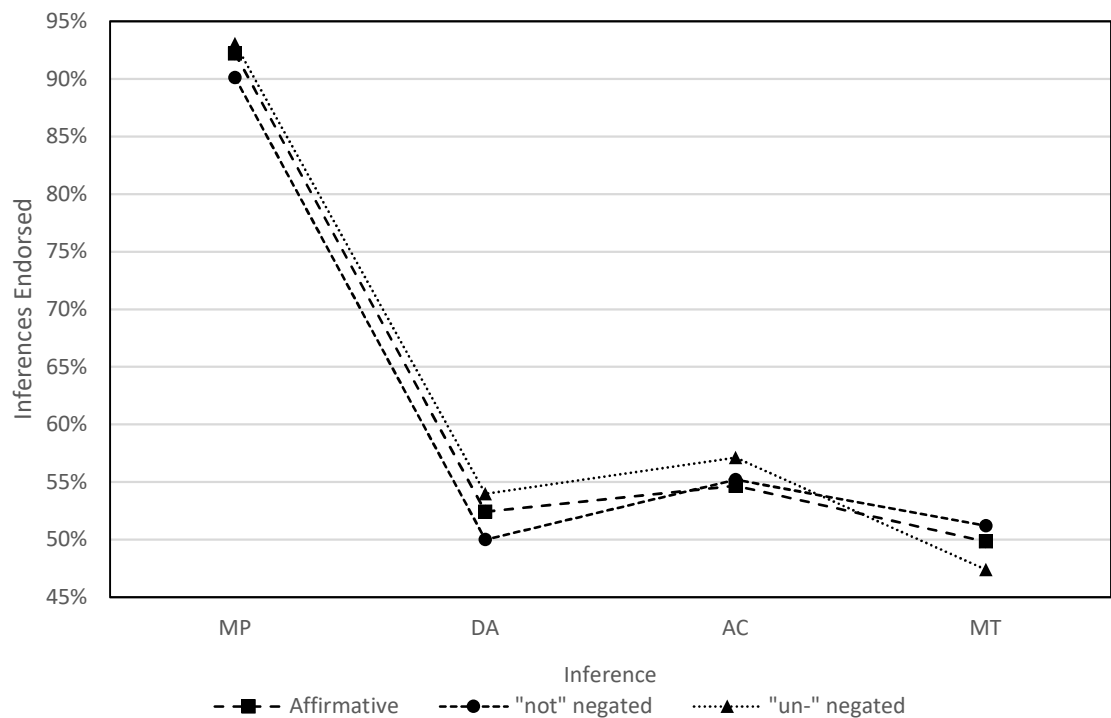
### 3.4.3. Results

Participants endorsed 62.3% of inferences overall in experimental tasks: 91.8% of MP inferences, 52.1% of DA, 55.7% of AC and 49.5% of MT. The proportion of inferences endorsed broken down by inference type and polarity of the inference conclusion is set out in table 3.4.

*Table 3.4 – Percentage of inferences endorsed in experiment 2*

Conclusion Polarity	Inference type			
	MP	DA	AC	MT
Affirmative	92.7%	52.0%	54.7%	49.3%
“not” negated	90.5%	52.4%	55.2%	49.8%
“un-” negated	92.2%	-	57.1%	-

A 4x3x3 ANOVA was conducted on the level of endorsement with inference type, antecedent polarity and consequent polarity as the factors and the responses to both scenarios combined. This found an interaction between inference type and antecedent polarity ( $F(6, 570) = 2.511, p=.021$ ) and a main effect of inference ( $F(3, 285) = 44.934, p<.001$ ). Comparison t-tests (with Bonferroni adjusted significance levels of .008) found that the main effect of inference represented a significant differences between MP inferences and DA inferences ( $t(95) = 8.625, p<.001$ ), between MP and AC ( $t(95) = 8.228, p<.001$ ) and between MP and MT ( $t(95) = 9.533, p<.001$ ). No significant differences were found between the DA, AC and MT. The interaction between inference type and antecedent polarity is shown in figure 3.4. Three comparisons were completed for each the four inference types (twelve comparisons in total). The compared responses with each antecedent polarity with each other antecedent polarity for a given inference level and used a Bonferroni adjusted significance level of  $p=.004$ . None of these comparisons met the required significant level. The only two comparisons that approached (but did not meet) the required significance level were between MP inferences with “not” negated and “un-” negated antecedents ( $t(95) = -2.674, p=.009$ ) and between DA inferences with “not” negated and “un-” negated antecedents ( $t(95) = -2.317, p=.023$ ).



*Figure 3.4* – Chart showing percentage of inferences by inference and antecedent polarity in experiment 2 (combined scenarios).

MP inferences with “un-” negated conclusions were slightly more endorsed than those with “not” negated conclusions. However, a one-way ANOVA on the proportion of MP responses endorsed with conclusion polarity as the factor (three levels: affirmative; “not” negated and “un-” negated) did not show a significant effect. Against the prediction that a negative conclusion bias would be observed, the MP inferences with affirmative conclusions were more endorsed than those with either form of negation in the conclusion.

AC inferences with “un-” negated conclusions were also more endorsed than those with “not” negated conclusions. However, no significant effect of conclusion polarity was found in one-way ANOVA on the proportion of AC responses endorsed with conclusion polarity as the factor (three levels: affirmative; “not” negated and “un-” negated).

Both DA and MT inferences showed a slightly higher level of responses when the conclusion was “not” negated than when it was affirmative. However, t-tests found that these differences were not significant.

The data were also fitted with Oaksford et al.'s (2000) computational-level model of conditional inference (see §1.3.3 for a description of this model). This model predicts inference levels for conditionals of the form 'if  $p$  then  $q$ ' based on the parameters  $P(p)$ ,  $P(q)$  and  $P(q | p)$  using the following formulae:

$$MP = P(q | p)$$

$$DA = (1 - P(q) - P(p) * (1 - P(q | p))) / (1 - P(p))$$

$$AC = (P(q | p) * P(p)) / P(q)$$

$$MT = (1 - P(q) - P(p) * (1 - P(q | p))) / (1 - P(q))$$

Fits were created for the combined scenarios and each scenario by setting each of the three parameters at 0.5 and then calculating the best fitting values for the other two for each polarity condition using the coefficient of variation as a measure of best fit for DA, AC and MT inferences. The best fits had all parameters set at a value of around 0.5 (the mean parameter values of the best fits for combined scenarios were  $P(p) = .52$ ,  $P(q) = .50$  and  $P(q | p) = .52$ ). The lowest coefficient of variation for these fits was  $R^2 = 0.93$  and most (76.2%) of the fits were  $R^2 > 0.99$  suggesting that the model was highly predictive of endorsement levels (the  $R^2$  statistic is described later in §4.2.4.1).

#### 3.4.3.1. Results for Each Scenario

To look at whether there were any scenario or context effects that could have obscured any potential findings, responses to each of the two scenarios will be considered separately.

When the first scenario was used, participants endorsed 61.6% of experimental task inferences overall. This was made up of 91.1% of MP, 52.3% of DA, 53.8% of AC and 49.1% of MT inferences. The proportions of scenario 1 inferences endorsed by inference type and conclusion polarity are shown in table 3.5.

*Table 3.5 – Percentage of inferences endorsed using scenario 1*

Conclusion Polarity	Inference type			
	MP	DA	AC	MT
Affirmative	92.4%	52.4%	52.8%	49.0%
“not” negated	89.9%	52.1%	53.8%	49.3%
“un-” negated	91.0%	-	54.9%	-

The level of endorsement when scenario 1 was used was tested using a 4x3x3 ANOVA with inference type, antecedent polarity and consequent polarity as factors. This found interactions between antecedent polarity and consequent polarity ( $F(4, 380) = 4.291$ ,  $p=.002$ ) and between inference type and antecedent polarity ( $F(6,570) = 2.208$ ,  $p=.041$ ). A main effect of inference ( $F(3, 285) = 41.974$ ,  $p<.001$ ) was also found. Comparison t-tests (with Bonferroni adjusted significance levels of  $p=.008$ ) found that the main effect of inference represented a significant difference between MP inferences and DA inferences ( $t(95) = 8.234$ ,  $p<.001$ ), between MP and AC ( $t(95) = 7.954$ ,  $p<.001$ ) and between MP and MT ( $t(95) = 9.183$ ,  $p<.001$ ). No significant differences were found between the DA, AC and MT inferences.

The interaction between inference type and antecedent polarity for scenario 1 is shown in figure 3.5. Twelve comparisons were completed, three for each the four inference types comparing responses with each antecedent polarity with each other antecedent polarity. No comparisons reached the Bonferroni adjusted significance level of  $p=.004$ . Two comparisons approach significance, which were between DA inferences with “not” negated and “un-” negated antecedents ( $t(95) = -2.400$ ,  $p=.018$ ) and between MT inferences with “not” negated and “un-” negated antecedents ( $t(95) = 2.222$ ,  $p=.029$ ).

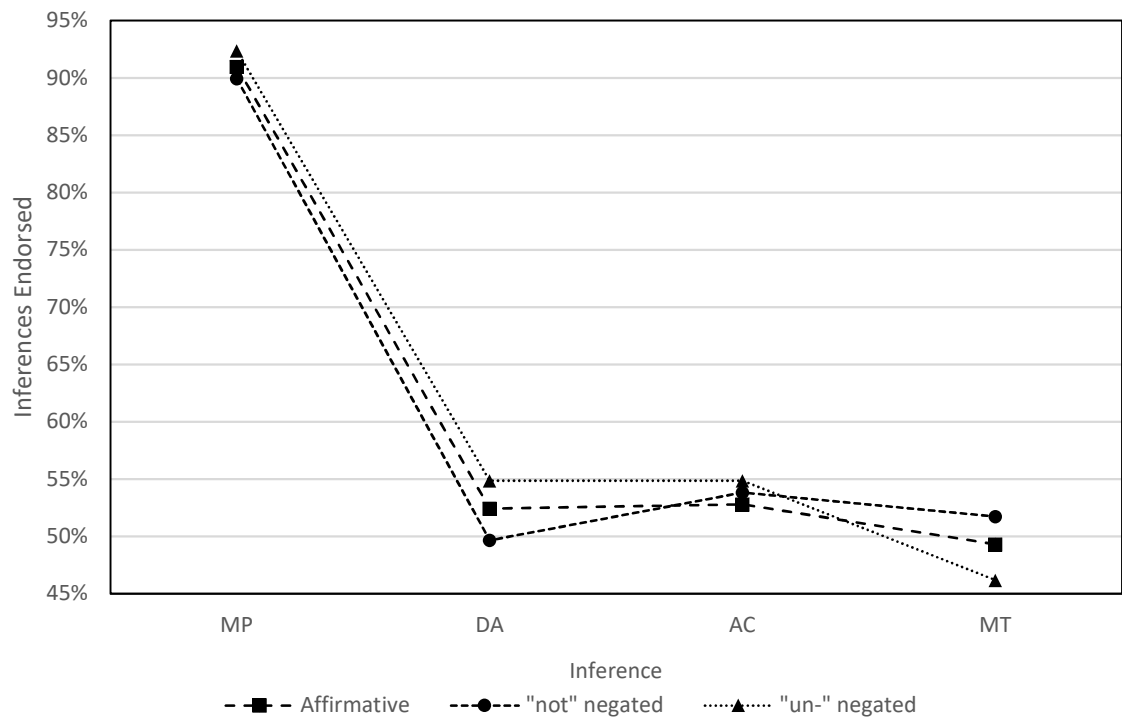
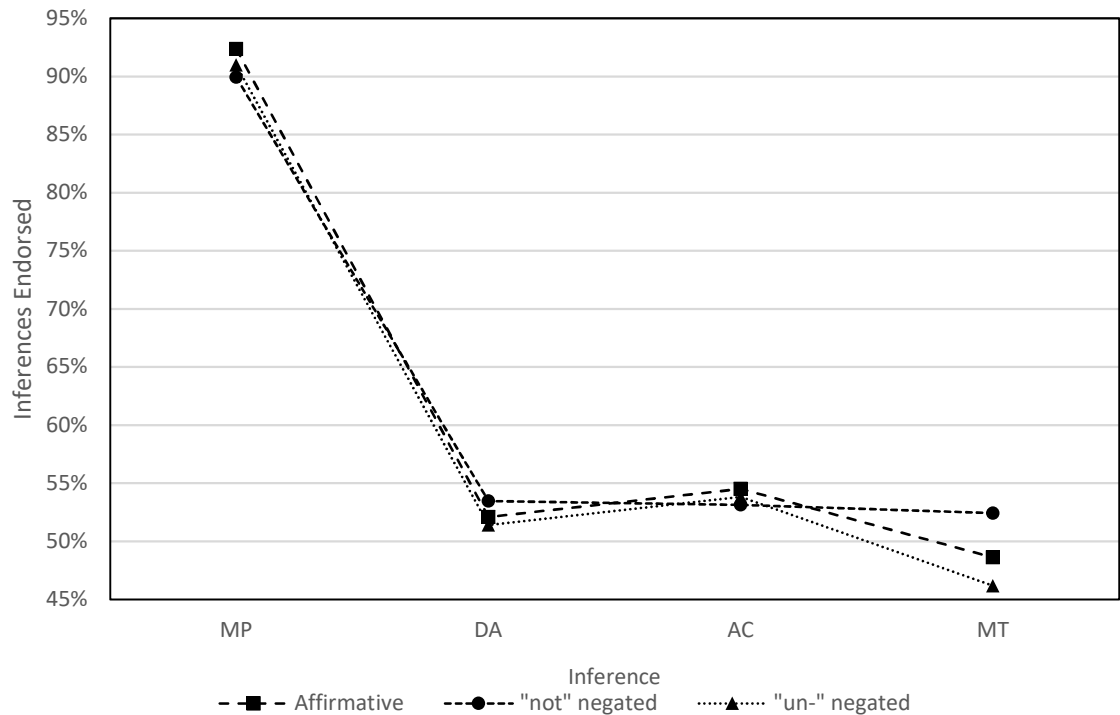


Figure 3.5 – Chart showing interaction between inference type and antecedent polarity (scenario 1).

The interaction found for scenario 1 between inference type and consequent polarity is shown in figure 3.6. Twelve comparisons were made, three for each the four inference types between responses with each consequent polarity with each other consequent polarity. No comparisons reached the Bonferroni adjusted significance level of  $p=.004$ . Only one comparison approached significance, which was between MT inferences with “not” negated and *un* negated antecedents ( $t(95) = 2.339$ ,  $p=.021$ ).



*Figure 3.6 – Chart showing interaction between inference type and consequent polarity (scenario 1).*

Participants endorsed a total of 62.96% of experimental task inferences when the second scenario was used. This was made up of 92.48% of MP, 51.97% of DA, 57.52% of AC and 49.88% of MT inferences. The proportions of inferences endorsed by inference type and conclusion polarity for scenario 2 are shown in table 3.6.

*Table 3.6 – Percentage of inferences endorsed using scenario 2*

Conclusion Polarity	Inference type			
	MP	DA	AC	MT
Affirmative	93.1%	51.6%	56.6%	49.7%
"not" negated	91.0%	52.8%	56.6%	50.4%
"un-" negated	93.4%	-	59.4%	-

Levels of endorsement for tasks using scenario 2 were compared with a 4x3x3 ANOVA with inference type, antecedent polarity and consequent polarity as factors. This found no interactions and a main effect of inference ( $F(3, 285) = 42.848, p < .001$ ) was also found. Comparison t-tests (with Bonferroni adjusted significance levels of  $p = .008$ ) found that the main effect of inference represented a significant differences between MP inferences and DA inferences ( $t(95) = 8.677, p < .001$ ), between MP and AC ( $t(95) =$

8.129,  $p < .001$ ) and between MP and MT ( $t(95) = 9.448$ ,  $p < .001$ ). No significant differences between the DA, AC and MT inferences were detected.

The pattern of responses for MP inferences was the same for scenario 1 as it was overall: inferences with affirmative conclusions more endorsed than those with “un-” negated conclusions; and inferences with “un-” negated conclusions were more endorsed than those with “not” negated conclusions. The pattern differed for scenario 2 with inferences that had “un-” negated conclusions more endorsed than those with affirmative conclusions (but both were still more endorsed than inferences with “not” negated conclusions). However, two one-way ANOVAs on the proportion of MP responses endorsed for each scenario with conclusion polarity as the factor (three levels: affirmative; “not” negated and “un-” negated) did not show any significant effects.

As with the combined results, both scenarios individually showed that AC inferences with “un-” negated conclusions were also more endorsed than those with “not” negated conclusions. Scenario 1, like the combined results, showed inferences with “not” negated conclusions were more endorsed than those with affirmative conclusions. Trials using scenario 2 showed that inferences with “not” negated and affirmative conclusions were endorsed at the same level. In neither case did the two scenarios show any significant effect of conclusion polarity on responses to AC inferences (in two one-way ANOVAs).

For MT inferences, both scenarios shared the overall pattern of results with inferences with conclusions using “not” negated slightly more endorsed than those with affirmative conclusions. However, for DA inferences, scenario 1 showed a higher level of endorsement where conclusions were affirmative than where they were “not” negated. For scenario 2, inferences with “not” negated conclusions were more endorsed than those with affirmative conclusions. A series of t-tests found that none of these small differences was significant.

#### **3.4.4. Discussion**

The results of this experiment did not support the hypothesis that using different types of negation in reasoning tasks would have different effects on people’s reasoning. No significant differences were found for inferences of each type when using different

polarity conditions (affirmative, “not” negation and “un-” negation) in their conclusions. When responses from both scenarios were combined, there was an interaction between antecedent polarity and inference type on the level of inferences endorsed. Looking just at scenario 1, there was also an interaction effect of antecedent polarity and inference type and an additional interaction effect of consequent polarity and inference type. However, it is hard to unpick these effects as none of the individual differences tested reached the required significance level and the pattern of results approaching significant was not consistent across both scenarios. It may be that other factors obscured any small effects.

Surprisingly, the results of this experiment did not show the effects of inference type or negative conclusion bias that have previously been widely observed. Most conditional inference evaluation tasks see different levels of response between DA, AC and MT inferences as well as between MP and the other forms of inference (e.g. Schroyens et al., 2001, discussed in §1.2). This experiment only found differences between MP and the other forms of inference. The other unusual result was the failure to detect any significant negative conclusion bias for any inference type. This raises two questions: why was the often seen differences between DA, AC and MT not observed; and why were levels of MP endorsement significantly higher than the other inferences despite whatever suppressed the normal differences between them?

As shown in the results, Oaksford et al.’s (2000) conditional probability model can provide a good fit for the data derived from this experiment. This model suggests that that endorsement of MP inferences using conditionals of the form ‘if  $p$  then  $q$ ’ is closely related to  $P(q \mid p)$ . Oaksford and Chater (2007, 2013) discuss the difference between MP inference levels and those of DA, AC and MT. They suggest that MP inferences are likely to be made without the same consideration of potential alternative causes and disabling propositions that we have seen incorporated into causal interpretations (e.g. Cummins et al., 1991) and which suppress inference levels. This provides a likely explanation for MP inference levels to remain high even when the plausibility of the conditional relationship is low (which would suppress other inference levels). In this case, as before, it looks like MP endorsement levels approached the ceiling level, potentially obscuring any small effects that might otherwise have been observed.



The normal differences between DA, AC and MT may have been suppressed because of an unusual characteristic of the scenarios. The two scenarios used were selected because of the lack of difference in the plausibility of the conditionals they used between different antecedent and consequent polarity conditions. This makes them quite unusual for a naturalistic scenario (people's normal expectation would reasonably be that 'if  $p$ , then  $q$ ' is highly plausible then 'if  $p$ , then not  $q$ ' would not be). To have this property, the conditions may have all been perceived as having ambiguous or uncertain plausibility in every polarity condition. Previous studies (e.g. Chan & Chau, 1994) have shown that the plausibility of premises given can affect people's willingness to draw inferences from them. This ambiguity may, therefore, have led participants to suppress the level of inferences people are willing to make or endorse using them. This suppression of inferences may have reduced the differences normally observed between the different forms of inference (and also suppressed any other negation effects that might have been observed).

The very good fit with Oaksford et al.'s (2000) computational-level model of conditional inference may provide further experimental support for the validity of that model and the probabilistic approach on which it was based. However, the overall lack of variability observed in the level of endorsement of DA, AC and MT means that the current data may be easily modelled. Therefore the current results do not provide a useful test for the model.

Although this experiment had substantially more participants than the previous experiment, it may not have been much more powerful overall. In the previous experiment, each of the nineteen participants responded to twelve different versions (one for each scenario) of each type of inference in each antecedent polarity condition – a total of 228 responses per inference and antecedent polarity form. The current experiment does have a much larger number of participants but they each only completed two trials for each type of inference with each antecedent and consequent polarity combination. The 96 participants therefore only provided 192 responses per inference and polarity form – a similar order of responses to the prior experiment. This undermines the attempt in this experiment to increase the size of the sample to help detect subtle effects.

Although these results do not support the hypothesis that different types of negation will influence inference rates in different ways, it looks like the scenarios used in this experiment did not elicit the responses that are normally seen. Therefore, it may be that the anticipated negation effects could still be observed using scenarios that provoked more typical responses. The plausibility of the conditionals used in the current scenarios may have been a factor in suppressing other inference effects and this highlights the importance of plausibility as a factor in inference judgements.

### **3.5. Experiment 3**

This experiment is intended to provide a further test of the differences that different types of negation would differently affect people's conditional inference judgements. Like the previous experiment, it uses the full extended negations paradigm (based on Evans', 1977, negations paradigm). As with the previous experiment, affirmative, "not" negated and "un-" negated antecedent and consequent terms are used in the conditional statements and MP, DA, AC and MT inferences are presented to participants.

The previous experiment attempted to use scenarios which minimised the variation in plausibility between the conditional statements used. This is an odd characteristic for conditional statements to have. It may have led to the conditional statement being perceived as having more ambiguous plausibility than a typical naturalistic conditional in all the antecedent and consequent polarity conditions. This unusual property may have contributed to the suppression of the effects of inference type normally observed in conditional inference validation tasks. The current experiment will, therefore, use scenarios that are expected to have varied plausibility across negation conditions – based on the scenarios used in chapter 2, experiment 3 (see §2.5.2.3). These conditionals have already been shown to have more natural plausibility characteristics, like being more plausible when both antecedent and consequent terms were affirmative or when both these terms were negative (either negation condition). The use of scenarios with more normal plausibility characteristics should facilitate more typical responses to the inference tasks than were observed in the previous experiment. Plausibility will, therefore, have to be considered as a covariant factor in the analysis of inference response.

In the prior experiment, all participants saw tasks based on the same scenario for 36 trials for each of the two experimental scenarios. This will have been quite repetitive for participants (even with distraction tasks randomly inserted). While the randomisation of trials should have avoided order effects, there is a risk that repetition led to participants making assumptions about what they were seeing and responding differently – or without paying sufficient attention – to later trials. Overall, this may have reduced the sensitivity of the tasks to the small effects that it was aiming to reveal. The current experiment, therefore, employs a design in which each participant only sees each scenario once. In the current experiment, each participant will only provide a single response for each scenario and each inference type. To retain the power to detect small effects, substantially more participants will need to be recruited than for the previous experiment.

### **3.5.1. Predictions**

As in the previous experiments, we expect to see a pattern of inference endorsement similar to that previously observed when affirmative and negated propositions are used (e.g. Oaksford et al., 2000). Where ‘un-’ negation is used, it is anticipated to behave more like affirmative than ‘not’ negated propositions with endorsement levels between those seen under those other two polarity conditions. This basic pattern may be obscured by any negative conclusion bias (e.g. Evans et al., 1995) which is anticipated to lead to higher levels of endorsement where conclusions are “not” negated than where they are “un-” negated or affirmative (“un-” negation is used only in some conclusions for MP and AC tasks). The plausibility of the conditional used will also be a significant factor (people will be more willing to endorse inferences where the conditional used has high plausibility in all inference conditions).

As set out before the prior two experiments, these predictions align with probabilistic computational level models (e.g. Oaksford et al., 2000) and CBN algorithmic models (e.g. Fernbach & Erb, 2013). They are not consistent with mental models accounts (e.g. Johnson-Laird & Byrne, 2002).

Ceiling effects may obscure any impact of using different types of negation in MP inferences.

### **3.5.2. Method**

#### **3.5.2.1. Design**

This experiment used inference type (four levels: MP, DA, AC and MT), antecedent polarity (three levels: affirmative, “not” negated and “un-” negated) and consequent polarity (the same three levels as antecedent polarity) as independent variables. The dependent variable is participants’ responses to the question of whether they agree with the given conclusion to the inference. This response was measured using an eight-point scale (from “Completely disagree” to “Completely agree”). Each participant would see a task for each inference type (each using a difference scenario) with polarity conditions randomised for each task.

The second part of the experiment uses antecedent polarity and consequent polarity as independent variables (each with the same three levels as above). The dependent variable is the probability rating assigned to the conditional statement provided (a number between 0 and 100). Participants would see the same four conditional statements that they had seen in the first part (where the polarity conditions had been randomly assigned).

#### **3.5.2.2. Participants**

Participants were recruited and paid through the ‘Amazon Mechanical Turk’ system (payment was US\$0.10). Potential duplicate responses (those from the same IP address or a Mechanical Turk ID that had been used in a prior experiment reported here) were removed. One participant was removed because their responses to the experimental survey were incompletely recorded. This left 848 participants with median age 27 (range 16 to 69). A total of 42% of participants were female and 50% had English as their first language. All completed the experiment online through [www.surveygizmo.com](http://www.surveygizmo.com).

#### **3.5.2.3. Materials**

Four scenarios were prepared to provide context for each conditional inference task. These each combined two of the scenarios used in chapter 2, experiment 3 (see §2.5.2.3). As before, these provided statements about the situation of 100 people and then further statements designed to qualitatively indicate that more people would share

each trait described than don't while providing ambiguity about the status of each individual. For example:

*100 people are asked if they believe in a newspaper's investigative report.*

- *Many of those people will be convinced by the evidence.*
- *A few of those people never trust reports.*
- *Several of those people normally accept such reports without much consideration.*
- *Some of those people will want to challenge the evidence.*

*The same 100 people are also asked what they plan for Valentines Day.*

- *Many of those people want to do something nice with their partners.*
- *A few of those people will upset their partners on Valentines Day.*
- *Several of those people spend a lot of time planning their evening.*
- *Some of those people are cynical about Valentines Day.*

For each of the scenarios, nine statements including a conditional sentence were prepared. These took a form like "An observer looks at the 100 people and suggests the following rule: If a person is believing, then the person is not romantic"). Nine forms of conditional statement were prepared (three with each antecedent polarity condition for each of the three consequent polarity conditions).

For the first part of the experiment, these scenarios were each presented with one of the conditional statements set out above as part of a conditional inference statement. Four conditional inference statements (one for each of the tested types of inference: MP, DA, AC and MT) were prepared for each of the nine conditional sentence forms prepared for each scenario (a total of 36 conditional inferences per scenario). As well as the scenario context and the statement including the conditional, each conditional inference statement included a premise related to the scenario and conditional sentence (e.g. "Given the rule, if a person from the 100 people was found meeting the following condition: The person is not romantic" for the MP case of the example given above) and

a conclusion (continuing the example: “Would you agree with the following conclusion? The person is believing”).

For the second part of the experiment, the same set of 36 conditional inference statements for each scenario were used without the premise and conclusion.

For more detail on the materials used, see appendix 2 (§A2.3).

#### 3.5.2.4. Procedure

Participants were recruited through the Mechanical Turk system and directed to the experimental survey. The first screen sought their informed consent and they were then provided with instructions for the inference tasks.

Each participant completed four inference tasks, one on each screen. The four tasks represented both one using each of the four scenarios and one using each of the inference types (MP, DA, AC and MT). The inference type combined with each scenario was randomly varied for each participant and the tasks were presented to each participant in random order. For each task, the antecedent polarity and consequent polarity conditions were randomised. This means that an individual participant may see more than one trial with the same combination of antecedent polarity and consequent polarity conditions. With only four trials per participant and nine possible combinations of antecedent polarity and consequent polarity conditions, no participant will see all of these conditions. The inference task will present the conditional inference statement as described in §3.5.2.3 above (including context, a statement with conditional sentence, premise and conclusion). Participants are asked to indicate whether they would endorse the inference by selecting one of the following options (presented in order) and clicking the ‘next’ button:

- Completely disagree
- Strongly disagree
- Disagree
- Disagree a little
- Agree a little
- Agree
- Strongly agree

- Completely agree

Once they had completed the four inference task questions, participants were given instructions for the second part of the experiment. In this part of the experiment, the participants were presented with the same conditional statements that they had seen in the first part of the experiment (with the context statements but without the premise and conclusion). They were then asked the following probability rating task question (Oaksford, Chater & Grainger, 1999): “For how many of the 100 people do you think the rule will be true for (enter a number between 0 and 100)?” This was intended to provide a measure of the plausibility assigned to the conditional statement by showing for how many people the participant thought the rule might be correct – more plausible rules expected to be perceived as correct for more people. Participants had to respond by typing a number between 0 and 100 and clicking next.

After the tasks were completed, participants were thanked for their time and invited to return to the Mechanical Turk system to enter a code for payment.

### **3.5.3. Results**

Responses to inference questions were recorded as a number between 1 and 8 with 1 corresponding to ‘Completely disagree’ with the inference and 8, ‘Completely agree’. The overall mean response score across all inference questions was 4.72 which would be placed between ‘Disagree a little’ and ‘Agree a little’ on the rating scale given – a little closer to ‘Agree a little’. For the different types of inference, the mean scores were 4.74 for MP, 4.81 for DA, 4.58 for AC and 4.75 for MT.

#### **3.5.3.1. Pattern of Inference Scores**

The mean and standard deviation of inference responses by inference type, antecedent polarity and consequent polarity are shown in table 3.7.

*Table 3.7 – Mean (standard deviation) inference scores by inference type, antecedent polarity and consequent polarity*

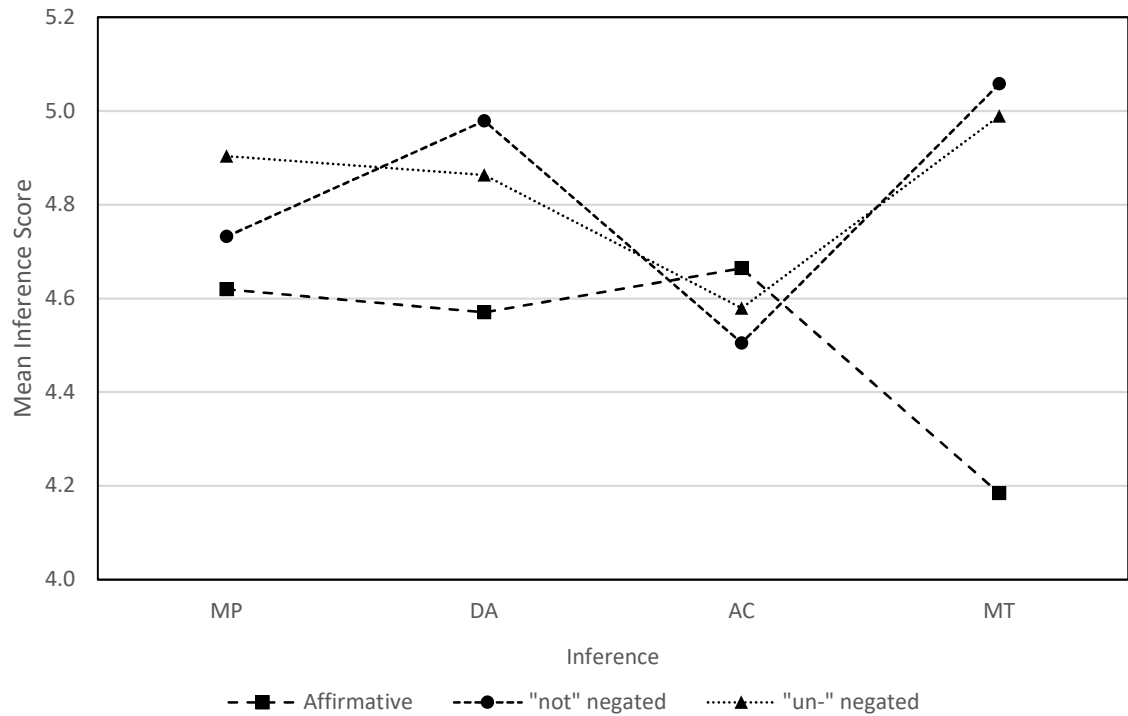
Inference Type	MP			DA		
	Antecedent Polarity					
Consequent Polarity	Affirmative	“not” negated	“un-” negated	Affirmative	“not” negated	“un-” negated
Affirmative	5.25 (1.68)	4.54 (1.71)	4.7 (1.68)	4.86 (1.64)	4.31 (2.00)	3.95 (1.99)
“not” negated	4.27 (1.96)	4.72 (1.81)	5.33 (1.85)	4.45 (1.94)	5.27 (1.49)	5.00 (1.81)
“un-” negated	4.35 (2.07)	4.95 (1.95)	4.7 (2.02)	4.33 (1.93)	5.36 (1.67)	5.49 (1.59)

Inference Type	AC			MT		
	Antecedent Polarity					
Consequent Polarity	Affirmative	“not” negated	“un-” negated	Affirmative	“not” negated	“un-” negated
Affirmative	5.46 (1.67)	4.32 (2.06)	4.1 (2.01)	4.5 (1.78)	4.25 (2.00)	4.23 (1.83)
“not” negated	4.06 (1.96)	4.69 (1.60)	4.63 (1.89)	4.14 (1.82)	5.33 (1.59)	5.22 (1.79)
“un-” negated	4.44 (1.69)	4.48 (1.68)	4.97 (1.96)	3.84 (2.01)	5.49 (1.71)	5.31 (1.82)

A 4x3x3 ANCOVA was used to compare inference response scores (each representing a participant’s level of agreement with a given inference) across inference types, antecedent polarities and consequent polarities. Plausibility of the relevant conditional used was treated as a covariate. With plausibility as a covariate removed, adjusted mean inference endorsement scores were 4.77 for MP, 4.78 for DA, 4.59 for AC and 4.70 for MT. This found an interactions between inference type and antecedent polarity ( $F(6, 3355) = 5.852, p < .001, \eta_p^2 = .010$ ) shown in figure 3.7, between inference type and consequent polarity ( $F(6, 3355) = 4.569, p < .001, \eta_p^2 = .008$ ) shown in figure 3.8 and between antecedent polarity and consequent polarity ( $F(4, 3355) = 11.477, p < .001, \eta_p^2 = .013$ ) shown in figure 3.9. It also found main effects of antecedent polarity ( $F(2, 3355)$



= 10.043,  $p < .001$ ,  $\eta_p^2 = .006$ ) and consequent polarity ( $F(2,3355) = 13.331$ ,  $p = .001$ ,  $\eta_p^2 = .008$ ). The covariate, plausibility ratings of the relevant conditionals, also had a significant effect on inference scores ( $F(1, 3355) = 490.308$ ,  $p < .001$ ,  $\eta_p^2 = .128$ ).



*Figure 3.7* – Chart showing interaction between inference type and antecedent polarity on inference responses in experiment 3.

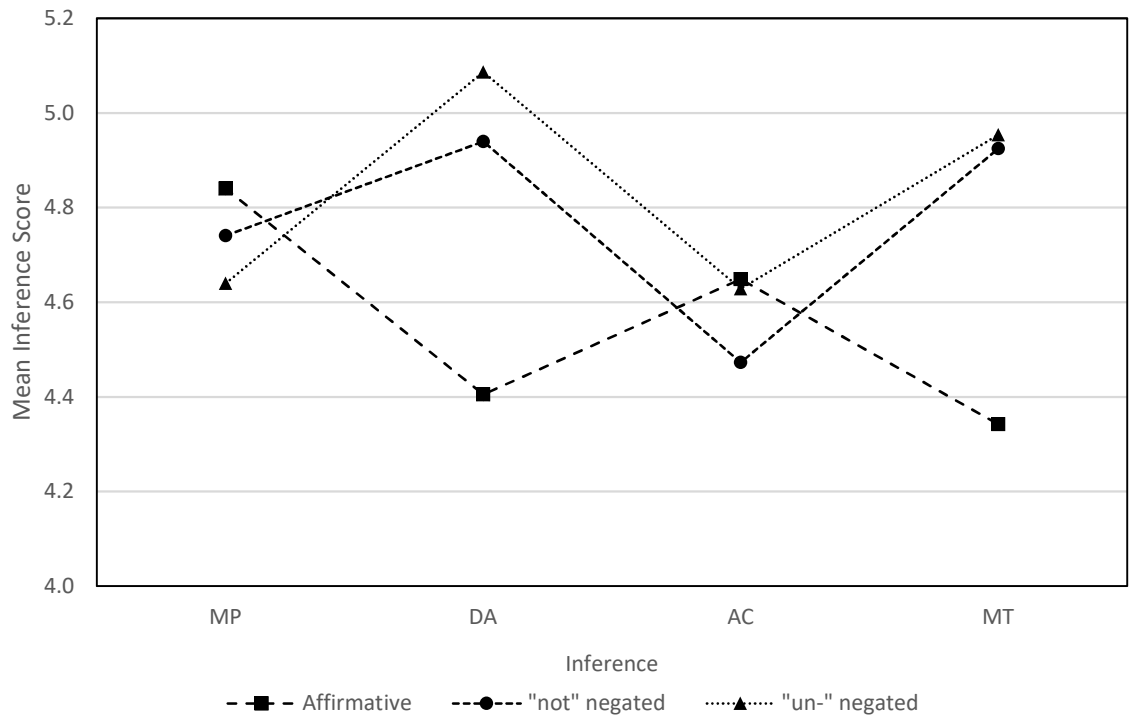


Figure 3.8 – Chart showing interaction between inference type and consequent polarity on inference responses in experiment 3.

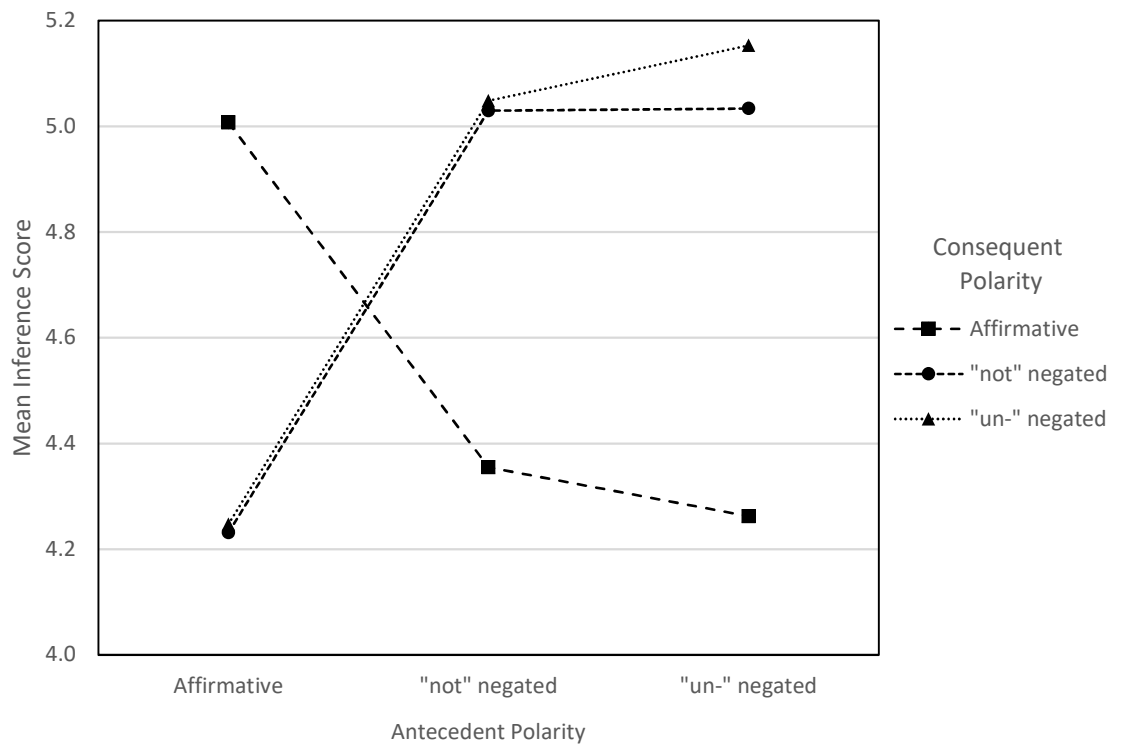


Figure 3.9 – Chart showing interaction between antecedent polarity and consequent polarity on inference responses in experiment 3.

Given the significant interactions, post hoc pairwise comparisons with Bonferroni adjustment for multiple comparisons were used to compare each condition for each of the three independent variables. For the inference conditions, no significant differences were found. For antecedent polarity, significant differences were observed between affirmative and “not” negated ( $p<.001$ ) and affirmative and “un-” negated ( $p<.001$ ) conditions. The mean response (adjusted for the covariate) was 4.52 when the antecedent was affirmative, 4.80 when it was “not” negated and 4.81 when it was “un-” negated. Similarly, for consequent polarity, significant differences were observed between affirmative and “not” negated ( $p=.001$ ) and affirmative and “un-” negated ( $p<.001$ ) conditions. The mean response (after adjustment) was 4.50 when the consequent was affirmative, 4.76 when it was “not” negated and 4.87 when it was “un-” negated.

### 3.5.3.2. Negative Conclusion Bias

The mean scores for each inference type and conclusion polarity is shown in table 3.8.

*Table 3.8 – Mean (standard deviation) responses to inference questions in chapter 3, experiment 3*

Conclusion Polarity	Inference type			
	MP	DA	AC	MT
Affirmative	4.84 (1.71)	5.01 (1.78)	4.66 (1.87)	5.02 (1.85)
“not” negated	4.74 (1.93)	4.40 (1.90)	4.50 (1.78)	4.18 (1.88)
“un-” negated	4.64 (2.03)	-	4.58 (1.98)	-

Two one-way ANCOVAs (with plausibility ratings covaried out) were conducted comparing inference response scores for MP and AC inferences with conclusion polarity as the factor. Neither ANCOVA found significant effects.

### 3.5.3.3. Role of Plausibility

The mean response to the probability rating tasks using the conditionals was 48.4 (standard deviation 27.7). The mean plausibility ratings given to each conditional with each antecedent and consequent polarity are shown in table 3.9.

*Table 3.9 – Mean plausibility ratings by antecedent polarity and consequent polarity*

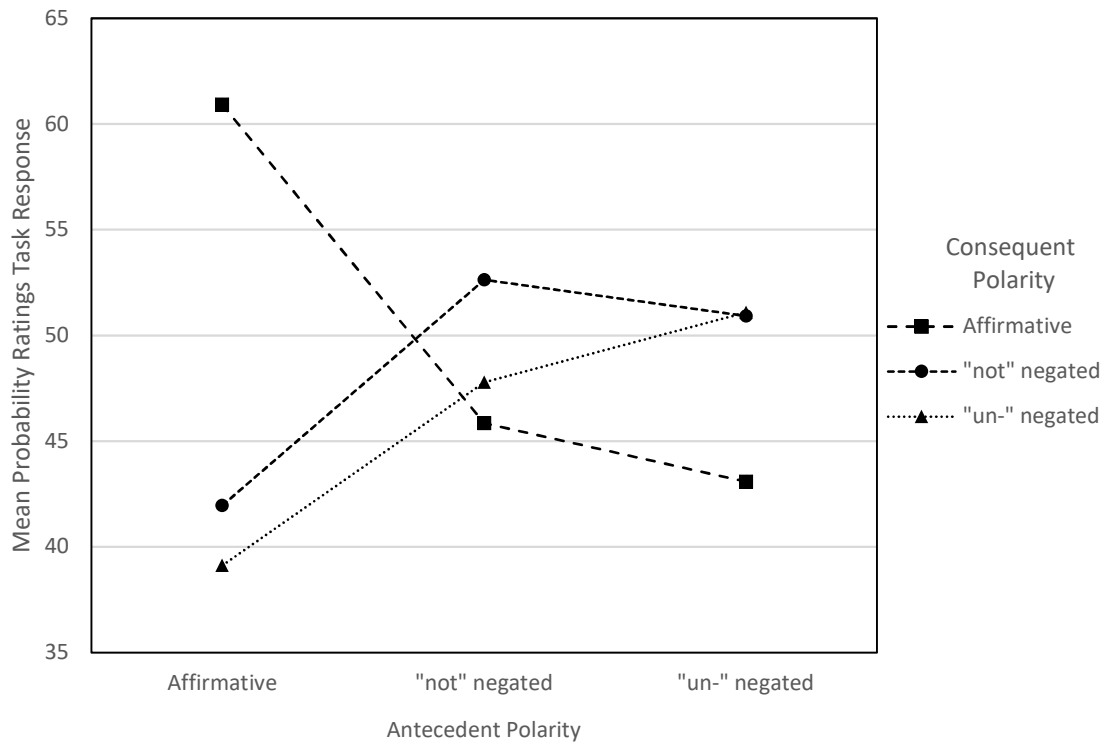
Consequent Polarity	Antecedent Polarity		
	Affirmative	“not” negated	“un-” negated
Affirmative	60.9	45.8	43.1
“not” negated	42.0	52.6	50.9
“un-” negated	39.1	47.8	51.1

Under probability theory, probabilities should meet the following constraint:

$$P(\neg q \mid p) + P(q \mid p) = 1$$

We can consider whether people’s plausibility ratings are consistent with this constraint for the probability of ‘if  $p$  then  $q$ ’ by looking at the sum of the plausibility ratings (divided by 100) for the affirmative consequent case and each of the cases with a negated consequent. On average these are 1.029 (for the “not” negated consequent case) and 1.000 (for the “un-” negated consequent case) where the antecedent is affirmative. Where antecedents are “not” negated, the equivalent values are .984 (for the “not” negated consequent case) and .936 (for the “un-” negated consequent case). Where antecedents are “un-” negated, the equivalent values are .940 (for the “not” negated consequent case) and .942 (for the “un-” negated consequent case). All of these results are close to 1, suggesting people’s average plausibility ratings provide perceived probabilities consistent with probability theory.

A 3x3 between subjects ANOVA was completed on probability rating tasks responses with antecedent polarity and consequent polarity as factors (each with three factors: affirmative, “not” negated and “un-” negated). This found a significant interaction between antecedent polarity and consequent polarity ( $F(4, 3383) = 42.621, p < .001$ ) and a significant main effect of consequent polarity ( $F(2, 3383) = 6.164, p = .002$ ). The interaction between antecedent and consequent polarity conditions is shown in figure 3.10. Post hoc Tukey HSD tests were used to compare each condition within both factors and a significant difference was found between probability rating responses where affirmative and “un-” negated consequent terms were used ( $p < .001$ ).



*Figure 3.10* – Chart showing interaction effect between antecedent and consequent polarity on probability ratings of conditionals in experiment 3.

There was a significant positive correlation ( $r = .375$ ,  $p < .001$ ) between the perceived plausibility of the conditional statement (measured using the probability rating task score) and the level of agreement with the inference using that statement (measured using the response to the inference task).

Two groups of responses were identified from within the responses based on the plausibility attributed to the conditional. There was a high plausibility group (991 responses where participants rated the plausibility of the conditionals as greater than or equal to 67 on the probability rating task) and a low plausibility group (1,108 responses with probability rating responses less than or equal to 33). The proportion of responses in each antecedent and consequent polarity condition allocated to each group is set out in table 3.10. The mean inference responses were 5.5 for the high plausibility group (standard deviation 1.69) and 3.91 for the low plausibility group (standard deviation 2.02). This difference was significant ( $t(2097) = 19.485$ ,  $p < .001$ ).

*Table 3.10 – Proportion of responses in each polarity condition allocated to each plausibility group*

Consequent Polarity	Antecedent Polarity		
	Affirmative	“not” negated	“un-” negated
Affirmative	45% / 15% (411)	29% / 36% (372)	23% / 39% (343)
“not” negated	22% / 43% (372)	31% / 24% (403)	31% / 31% (357)
“un-” negated	17% / 48% (368)	28% / 33% (374)	33% / 29% (392)

Percentage of responses allocated to high plausibility group / Percentage of responses allocated to low plausibility group (Total number of responses)

Looking just at the high plausibility group, a 4x3x3 between participants ANOVA with inference type, antecedent polarity and consequent polarity as factors compared responses to the inference tasks. This showed an interaction effect between antecedent polarity and consequent polarity ( $F(4,955) = 3.877$ ,  $p=.004$ ,  $\eta_p^2 = .016$ ). Inference scores were higher where antecedent and consequent polarities are either both affirmative or both negative (i.e. using either negation condition) than when one polarity was affirmative and the other negative. Post hoc Tukey HSD tests were used to compare each condition within these two factors, but no significant differences were found.

Comparing only responses to inference tasks allocated to the low plausibility group, a 4x3x3 between participants ANOVA was conducted with inference type, antecedent polarity and consequent polarity as factors. This showed interactions between all three factors ( $F(12,1072) = 2.089$ ,  $p=.015$ ,  $\eta_p^2 = .023$ ). It also showed interactions between antecedent polarity and consequent polarity ( $F(4,1072) = 4.096$ ,  $p=.003$ ,  $\eta_p^2 = .015$ ) and between antecedent polarity and inference ( $F(6,1072) = 2.884$ ,  $p=.009$ ,  $\eta_p^2 = .016$ ). Where consequent polarity was negative, inference scores were higher when antecedent polarities were also negative than when they were affirmative but where consequent polarity was affirmative the opposite pattern was observed. Inference scores were higher when antecedent polarity was negative than when it was positive for MP, DA and MT inferences but this pattern was reversed for AC inferences. Main effects of antecedent polarity ( $F(2,1072) = 7.969$ ,  $p<.001$ ,  $\eta_p^2 = .015$ ) and consequent polarity ( $F(2,1072) = 6.212$ ,  $p=.002$ ,  $\eta_p^2 = .011$ ) were observed. Both polarity factors typically led to higher inference scores where they were negative. Post hoc Tukey HSD tests were used to compare each condition within these three factors involved in significant interactions.

For antecedent polarity, significant differences were observed between the affirmative condition and both “not” negation ( $p < .001$ ) and “un-” negation conditions ( $p < .001$ ). For consequent polarity, significant differences were observed between the affirmative condition and both “not” negation ( $p = .009$ ) and “un-” negation conditions ( $p = .010$ ).

#### 3.5.3.4. Other Possible Factors

Several other factors may have influenced responses to the inference tasks in this experiment. The high level of participants with English as a second language, the potential for participant or scenario effects and the fact that responses to inference tasks in this experiment were based on a rating scale rather than categorical as in previous experiments should all be considered.

A large proportion of participants did not have English as a first language. A 4x3x3 between participants ANCOVA was used to compare inference scores across inference types, antecedent polarities and consequent polarities for only those participants that reported English as a first language. Plausibility ratings were used as covariates. As with the same test using all participants in §3.5.3.1, this found interactions between inference type and antecedent polarity ( $F(6,1675) = 4.043$ ,  $p < .001$ ,  $\eta_p^2 = .014$ ), between inference type and consequent polarity ( $F(6,1676) = 2.280$ ,  $p = .034$ ,  $\eta_p^2 = .008$ ) and between antecedent polarity and consequent polarity ( $F(4,1676) = 5.145$ ,  $p < .001$ ,  $\eta_p^2 = .012$ ). It also found a main effects of antecedent polarity ( $F(2,1676) = 9.190$ ,  $p < .001$ ,  $\eta_p^2 = .011$ ) and consequent polarity ( $F(2,1676) = 8.008$ ,  $p < .001$ ,  $\eta_p^2 = .009$ ). Post hoc pairwise comparisons (with Bonferroni adjustment for multiple comparisons) comparing each condition for each of the three independent variables only showed any significant results for consequent polarity. Here significant differences were observed between affirmative and “un-” negated conditions ( $p < .001$ ). The mean response (after adjustment were) was 4.59 when the consequent was affirmative, 4.81 when it was “not” negated and 5.01 when it was “un-” negated.

For analysis purposes, this has been treated as a between participants design as all participants responded to different conditions. However, each participant gave responses to four inference tasks which may not be independent. A 4x3x3 between participants ANCOVA with inference types, antecedent polarity and consequent

polarity as factors and plausibility rating and participant as covariates found no significant effect of the participant covariate ( $F(1,3354) = 0.025, p=.876$ ).

This experiment used four scenarios. A  $4 \times 4 \times 3 \times 3$  ANCOVA compared inference scores with inference types, scenario, antecedent polarity and consequent polarity as factors (with plausibility ratings as covariate). This found a significant interaction between scenario, antecedent polarity and consequent polarity ( $F(12,3247) = 8.871, p<.001, \eta_p^2 = .011$ ). It also found a main effect of scenario ( $F(3,3247) = 3.610, p=.013, \eta_p^2 = .003$ ). Post hoc pairwise comparisons (with Bonferroni adjustment) were used to compare each of the four scenario conditions with each other and significant differences were found between scenarios 3 and 4 ( $p=.007$ ). This appears to be a reflection of lower levels of inference endorsement when scenario 4 was used with negative antecedent and consequent polarity conditions than when scenario 3 was used in the same polarity conditions. Effects similar to those observed when scenario was not included as a factor were found between antecedent polarity and consequent polarity ( $F(4,3248) = 10.650, p<.001, \eta_p^2 = .013$ ), between inference and antecedent polarity ( $F(6,3248) = 5.816, p<.001, \eta_p^2 = .011$ ), between inference and consequent polarity ( $F(6,3248) = 4.623, p<.001, \eta_p^2 = .008$ ) and main effects of antecedent polarity ( $F(2,3248) = 8.439, p<.001, \eta_p^2 = .005$ ) and consequent polarity ( $F(2,3248) = 12.793, p<.001, \eta_p^2 = .008$ ).

This experiment sought responses on an eight-point scale where the previous two experiments in this chapter had sought categorical responses. To look at whether this might have affected the findings, responses on the eight-point scale were converted to binary responses with scores of 1 to 4 treated as reflecting that the inference would not be endorsed and 5 to 8 as representing endorsement of the inference. Under this measure, endorsement levels were 59.7% for MP, 61.8% for DA, 56.5% for AC and 59.6% for MT inferences.  $4 \times 3 \times 3$  between participants ANCOVA was used to compare binary inference scores across inference types, antecedent polarities and consequent polarities (with plausibility rating as a covariate). This found an interactions between inference type and antecedent polarity ( $F(6,3355) = 4.708, p<.001, \eta_p^2 = .008$ ), between inference type and consequent polarity ( $F(6,3355) = 6.057, p<.001, \eta_p^2 = .011$ ) and between antecedent polarity and consequent polarity ( $F(4,3355) = 9.614, p<.001, \eta_p^2 = .011$ ). It also found a main effects of antecedent polarity ( $F(2,3355) = 5.892, p=.003, \eta_p^2 = .004$ ) and consequent polarity ( $F(2,3355) = 11.110, p<.001, \eta_p^2 = .007$ ).



### 3.5.4. Discussion

The pattern of results in this experiment showed some unexpected characteristics. Firstly, inference levels did not align with results from prior experiments where Evans' (1977) negations paradigm has been used (e.g. Oaksford et al., 2000). Figure 3.11 shows how the responses here look if the different types of negation are reduced to the standard negation paradigm. MP normally sees the highest inference levels but, in this experiment, overall MP inference levels were lower than DA and MT (as the figure shows MP was only highest where antecedents were negated and consequents affirmative). AC is also typically higher than DA, but this standard result was also reversed overall (an effect that appears to be driven where consequents are negated).

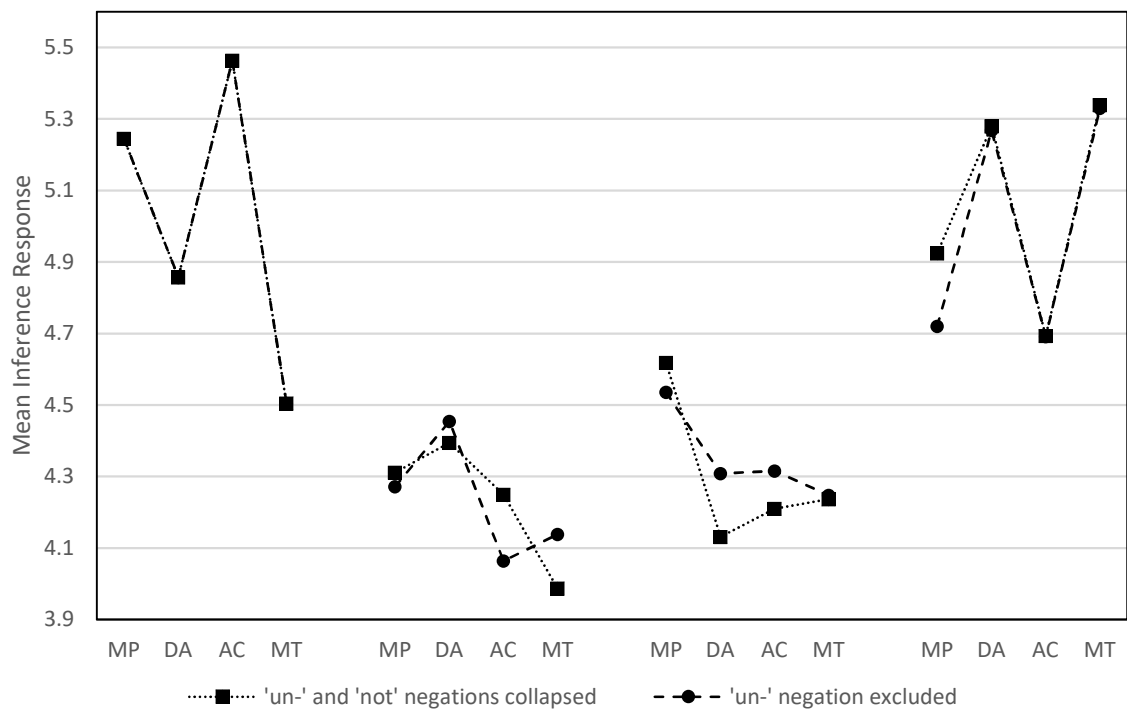


Figure 3.11 – Chart showing mean inference responses in experiment 3 by inference type and standard negation paradigm conditional formats.

The panels show (from left to right): (i) responses where conditionals had affirmative antecedents and affirmative consequents; (ii) responses where conditionals had affirmative antecedents and negated consequents; (iii) responses where conditionals had negated antecedents and affirmative consequents; and (iv) responses where conditionals had negated antecedents and negated consequents. Different lines show inference responses with the results from conditionals using “un-” negation excluded and with responses to conditionals using the two different types of negation averaged.

In contrast to the anticipated effect, where differences between polarity conditions were observed, it generally showed that the effects of “un-” negated propositions was closer to the effects of “not” negated propositions than the effect of affirmative propositions.

Also, no evidence of the typical negative conclusion bias (e.g. Evans et al., 1995) was seen – negated conclusions all saw lower levels of inference endorsement where conclusions used either type of negation than where conclusions were affirmative (although these differences were not significant).

This suggests that the way participants treated these scenarios is different to respondents to typical inference tasks – as with the prior experiments. For example, it has previously been seen that, where naturalistic scenarios are used, different factors can have greater influence on the results than is normally observed (Oaksford & Stenning, 1992).

These results are obviously not consistent with the expectations that responses would align with previous patterns of results with “un-” negation behaving more like affirmation than “not” negation. These predictions were based on probabilistic and CBN models. However, it is not clear that other approaches might provide an adequate explanation of these results. In particular, the mental models account (e.g. Johnson-Laird & Byrne, 2002) would not appear to be able to explain the lower levels of MP inference (which should normally be well facilitated by the initial mental model).

What the results did show, which was consistent with probabilistic and CBN accounts, was a highly significant positive correlation between the plausibility of the conditional statement (as measured through probability ratings task responses) and level of inference endorsement (measured through responses to inference tasks). This can be seen, particularly by looking at the suppression of inference levels where conditionals had incongruent antecedent and consequent polarities (which yielded lower plausibility ratings for the conditional). This may suggest that – for these scenarios – the use of ‘un-’ negation does yield similar consideration of enablers and disablers as ‘not’ negation. This may be consistent with the relatively wide scope interpretations “un-” negated statements can take compared to other forms of affixal negation like “in-” (Zimmer, 1964).

The impact of plausibility may also explain the main effect of scenario observed. Participants may have viewed scenario 4 statements like “If a person is not stylish, then the person is not wise” (mean plausibility score 46.31) as having less related antecedent and consequent terms than the equivalent statements in scenario 2 (“If a person is not friendly, then the person is not social”, mean plausibility score 58.70) and scenario 3 (“If a person is not grateful, then the person is not happy” mean plausibility score 51.79). Where there is perceived to be a limited connection between antecedent and consequent terms the conditional is often considered less plausible (e.g. Quinn & Markovits, 1998). Participants appear to have been more willing to draw inference when presented with the conditionals rated as more plausible. Such material effects may also explain why MP inferences, in particular, may trigger more consideration of disablers (alternatives to the antecedents that may cause the consequent). Under Fernbach and Erb’s CBN model for MP, this would reduce MP inferences.

While the results did not reflect the detail of the predictions made, this strong relationship between plausibility and inference levels does provide strong support for probabilistic and CBN approaches. As discussed above, it is likely that particular material effects changed the way these have operated in other experiments leading to the results observed.

Another possible cause of difference is that this experiment used a different measure to the previous experiments in this chapter. This measure – an eight-point scale – may be interpreted and treated differently by participants to a straightforward endorsement question requiring a binary response. When the inference responses on an eight-point scale were converted to a categorical measure, the pattern of results did not change. However, it remains possible that these results are not comparable to those in the prior experiments because of the measure used. It is possible that when faced with a wider range of optional responses, participants tend towards the middle of that range if they have any uncertainty. In the case of MP inferences, the option to say they ‘Disagree a little’ with the inference may be easier to select if the participant is uncertain than a binary response rejecting the inference completely.

The pattern of results remained the same when looking exclusively at participants who had English as a first language. This suggested that the large number of respondents

with English as a second language was not a significant factor affecting responses to the inference tasks.

This overall effect of plausibility, in an experiment which used naturalistic materials from which people can rate plausibility with reference to external knowledge, may have masked any other effects of inference. This effect may also have outweighed the small effects when the type of negation was varied that had been hypothesised. It had been hoped that a much larger sample in an experiment (where the conditional statements would have more natural plausibility functions than in experiment 2) might have helped the observation of small negation effects. There were substantially more respondents to this experiment than the previous two experiments in this chapter. However, the numbers were not as high as had been hoped. The design used meant that each participant only answered questions covering four of the 36 possible combinations of antecedent polarity, consequent polarity and inference type. This meant that, although there were 849 participants (almost nine times as many as in experiment 2 and almost 45 times the number in experiment 1), there were only an average of 94 responses for each combination of antecedent and consequent polarity and inference type (roughly half the number of responses per condition seen in the prior two experiments). This substantially undermines the ability of this experiment to detect subtle effects compared to the prior experiments.

The results of this experiment, therefore, have not allowed a more powerful test of predictions about the use of different types of negation in conditional inference tasks effectively. However, it has provided substantial further support for the hypothesis that plausibility is important to people's responses to inference tasks. It showed that people were significantly more likely to endorse an inference where the conditional used was rated as more plausible.

### **3.6. Experiment 4**

The previous experiment aimed to use more participants and a design in which each participant did not see each polarity condition to detect small effects caused by the use of different types of negation. However, fewer responses per combination of polarity and inference conditions were obtained than in the prior experiments. Also, the quasi-between-participants nature of the previous experiment (each participant completed

tasks in a random subset of conditions) prevented a fully crossed analysis of responses. These factors will have made it less sensitive than the prior experiments to small effects.

This experiment, therefore, aims to provide a more sensitive test of the effects of different types of negation on conditional reasoning. Like the previous two experiments in this section, this experiment will employ the full extended negations paradigm. Affirmative, “not” negated and “un-” negated antecedent and consequent terms will be used in the conditional statements for MP, DA, AC and MT inferences.

The current experiment will use the same scenarios as the prior experiment in a fully within-participants design. This requires each participant to respond to 144 inference questions (and a large number of plausibility questions). Around 50 participants would produce about 200 responses for each of the 36 combinations of inference and polarity conditions – similar to the number in the first two experiments in this section (and about twice the number in the third. The current experiment will, therefore, seek responses from more than 50 participants to increase the number of responses per condition. There is a risk of participants losing engagement and abandoning the experimental survey which will be avoided by increasing the payment available for participation.

As in the prior experiment, participants will be asked to rate the plausibility of the conditionals used after the experimental inference questions. In addition to considering the conditionals, participants will be asked to provide probability ratings for the antecedent propositions used in the conditions (in all three polarity conditions). This will enable easier comparison between participant responses and conditional probability models (e.g. Oaksford et al., 2000).

### **3.6.1. Predictions**

The predictions of this experiment are essentially the same as those set out for the previous experiment (see §3.5.1). As before, it is anticipated that the pattern of inference endorsement will be similar to that normally observed when affirmative and negated propositions are used (e.g. Oaksford, Chater & Larkin, 2000). Negative conclusion bias (in which conclusions using “not” negation are more likely to be endorsed, e.g. Evans et al., 1995) is also anticipated. Within both of these predicted effects, where ‘un-’ negation is used in propositions and conclusions, it is anticipated to

behave more like equivalent affirmative than ‘not’ negated propositions or conclusions. The plausibility of the conditional used is expected to be a significant factor and any differences in MP inference levels are likely to be obscured by ceiling effects.

As before, these predictions are based on probabilistic (e.g. Oaksford et al., 2000) and CBN (e.g. Fernbach & Erb, 2013) accounts but not mental models accounts (e.g. Johnson-Laird & Byrne, 2002).

### **3.6.2. Method**

#### **3.6.2.1. Design**

This experiment used a within-participants design with inference type (four levels: MP, DA, AC and MT), antecedent polarity (three levels: affirmative, “not” negated and “un-” negated) and consequent polarity (the same three levels as antecedent polarity) as independent variables. The dependent variable is participants’ responses to the question of whether they agree with the given conclusion to the inference. This response was measured using an eight-point scale (from “Completely disagree” to “Completely agree”).

The second part of the experiment will also be a within-participants design. The dependent variable is the probability rating assigned to the statement provided (a number between 0 and 100). The independent variable will be the statement, with 48 levels. The statements are made up of two forms. Firstly, the conditionals used (with antecedent and consequent negation systematically varied by the three polarity conditions – affirmative, “not” negated and “un-” negated – to provide nine conditionals for each of four scenarios). Secondly, statements of the antecedent term in each of the three polarity conditions (affirmative, “not” negated and “un-” negated to provide a further three statements for each of the four scenarios).

#### **3.6.2.2. Participants**

Participants were recruited and paid through the ‘Prolific Academic’ system (payment was £6.00). The system is similar to the Mechanical Turk system used in other experiments. Potential duplicate responses (those from the same IP address as another response) and incomplete responses were removed. This left 89 participants with median age 34 (range 20 to 62). A total of 48% of participants were female and 99%

had English as their first language. All completed the experiment online through [www.surveygizmo.com](http://www.surveygizmo.com).

### 3.6.2.3. Materials

This experiment used the same four scenarios that were prepared for the previous experiment (see §3.5.2.3). As before, nine statements including a conditional sentence were used for each scenario (three with each antecedent polarity condition for each of the three consequent polarity conditions). For each conditional statement, four inference tasks were again used (MP, DA, AC, MT). In this experiment, all participants would do all 144 tasks prepared rather than a subset of them (as was the case in the previous experiment).

As with the prior experiment, the scenarios and conditional statements (nine for each of the four scenarios) were used. Also, for this second part of the experiment people were asked to evaluate the plausibility of a further three statements for each scenario. These were statements that the antecedent term of the conditional was true in each polarity condition (affirmative, “not” negated and “un-” negated). This meant a total of twelve statements would be evaluated for each of the four scenarios (a total of 48).

For more detail on the materials used, see appendix 2 (§A2.3).

### 3.6.2.4. Procedure

Participants were recruited through the ‘Prolific Academic’ system and directed to the experimental survey. The first screen sought their informed consent and they were then provided with instructions for the inference tasks.

Each participant completed all 144 inference tasks, one on each screen, in random order. Each inference task presented the conditional inference statement as described above (including context, a statement with conditional sentence, premise and conclusion). Participants are asked to indicate whether they would endorse the inference by selecting one of the following options (presented in order) and clicking the ‘next’ button:

- Completely disagree
- Strongly disagree
- Disagree

- Disagree a little
- Agree a little
- Agree
- Strongly agree
- Completely agree

Once they had completed the inference task questions, participants were given instructions for the second part of the experiment. In this part of the experiment, the participants were presented with each conditional statement that they had seen in the first part of the experiment (with the context statements but without the premise and conclusion) and each antecedent proposition (in each polarity condition). These were provided one to a page in random order. They were then asked the probability ratings task question (Oaksford, Chater & Grainger, 1999): “For how many of the 100 people do you think the [rule or statement] will be true for (enter a number between 0 and 100)?” This was intended to provide a measure of the plausibility assigned to the conditional statement by showing for how many people the participant thought the rule might be correct – more plausible rules expected to be perceived as correct for more people. Participants had to respond by typing a number between 0 and 100 and clicking next.

After the tasks were completed, participants were thanked for their time and invited to return to the Prolific Academic system to enter a code for payment.

### **3.6.3. Results**

As with the previous experiment, responses to inference questions were recorded as a number between 1 and 8 with 1 corresponding to ‘Completely disagree’ with the inference and 8, ‘Completely agree’. The overall mean response score across all inference questions was 4.89 which would be placed between ‘Disagree a little’ and ‘Agree a little’ on the rating scale given – much closer to ‘Agree a little’. For the different types of inference, the mean scores were 5.25 for MP, 4.77 for DA, 4.84 for AC and 4.73 for MT.



### 3.6.3.1. Pattern of Inference Scores

The mean and standard deviation of inference responses by inference type, antecedent polarity and consequent polarity are shown in table 3.11.

*Table 3.11 – Mean (standard deviation) inference scores by inference type, antecedent polarity and consequent polarity*

Inference Type	MP			DA		
	Antecedent Polarity					
Consequent Polarity	Affirmative	“not” negated	“un-” negated	Affirmative	“not” negated	“un-” negated
Affirmative	6.1 (1.21)	4.64 (1.85)	4.65 (1.88)	5.08 (1.42)	3.89 (1.63)	3.92 (1.6)
“not” negated	4.56 (1.92)	5.7 (1.37)	5.69 (1.33)	3.95 (1.56)	5.55 (1.39)	5.49 (1.42)
“un-” negated	4.58 (1.92)	5.69 (1.32)	5.64 (1.34)	3.95 (1.55)	5.53 (1.39)	5.54 (1.44)

Inference Type	AC			MT		
	Antecedent Polarity					
Consequent Polarity	Affirmative	“not” negated	“un-” negated	Affirmative	“not” negated	“un-” negated
Affirmative	5.75 (1.28)	4.15 (1.58)	4.19 (1.7)	5 (1.43)	4.01 (1.5)	4.03 (1.58)
“not” negated	4.24 (1.55)	5.28 (1.36)	5.22 (1.34)	3.84 (1.62)	5.47 (1.5)	5.43 (1.41)
“un-” negated	4.18 (1.62)	5.28 (1.43)	5.23 (1.39)	3.9 (1.63)	5.43 (1.46)	5.48 (1.47)

For each participant, inference ratings and plausibility ratings for each experimental condition were averaged across the four scenarios. A 4x3x3 within participants Multivariate ANOVA (MANOVA) was used to compare mean inference response scores and mean conditional plausibility scores as dependent variables from each participant across the factors inference type (MP, DA, AC, MT), antecedent polarity and consequent polarities (the latter two both have three levels: affirmative, “not” negated, “un-” negated). This found significant multivariate effects for the interactions between inference type and antecedent polarity ( $F(6, 528) = 18.034, p < .001, \eta_p^2 = .170$ ), between

inference type and consequent polarity ( $F(6, 528) = 19.001, p < .001, \eta_p^2 = .178$ ) and between antecedent polarity and consequent polarity ( $F(8, 704) = 45.747, p < .001, \eta_p^2 = .342$ ). Multivariate analysis also found main effects of inference type ( $F(3, 264) = 9.190, p < .001, \eta_p^2 = .095$ ), antecedent polarity ( $F(4, 352) = 32.625, p < .001, \eta_p^2 = .270$ ) and consequent polarity ( $F(4, 352) = 29.703, p < .001, \eta_p^2 = .252$ ).

Univariate analysis for the effects of the independent variables on inference ratings found interactions between inference type and antecedent polarity ( $F(6, 1056) = 18.034, p < .001, \eta_p^2 = .170$ ) shown in figure 3.12, between inference type and consequent polarity ( $F(6, 1056) = 19.001, p < .001, \eta_p^2 = .178$ ) shown in figure 3.13 and between antecedent polarity and consequent polarity ( $F(4, 1056) = 101.377, p < .001, \eta_p^2 = .535$ ) shown in figure 3.14. It also found main effects of inference type ( $F(3, 1056) = 9.190, p < .001, \eta_p^2 = .095$ ), antecedent polarity ( $F(2, 1056) = 81.526, p < .001, \eta_p^2 = .481$ ) and consequent polarity ( $F(2, 1056) = 71.204, p < .001, \eta_p^2 = .447$ ). Note that the univariate effects related to inference are the same as those described in the multivariate effects above. This is because plausibility ratings were the same across inference types and so the effect of inference type is entirely on inference responses. Univariate analysis for the effects of the independent variables on plausibility ratings will be discussed below.

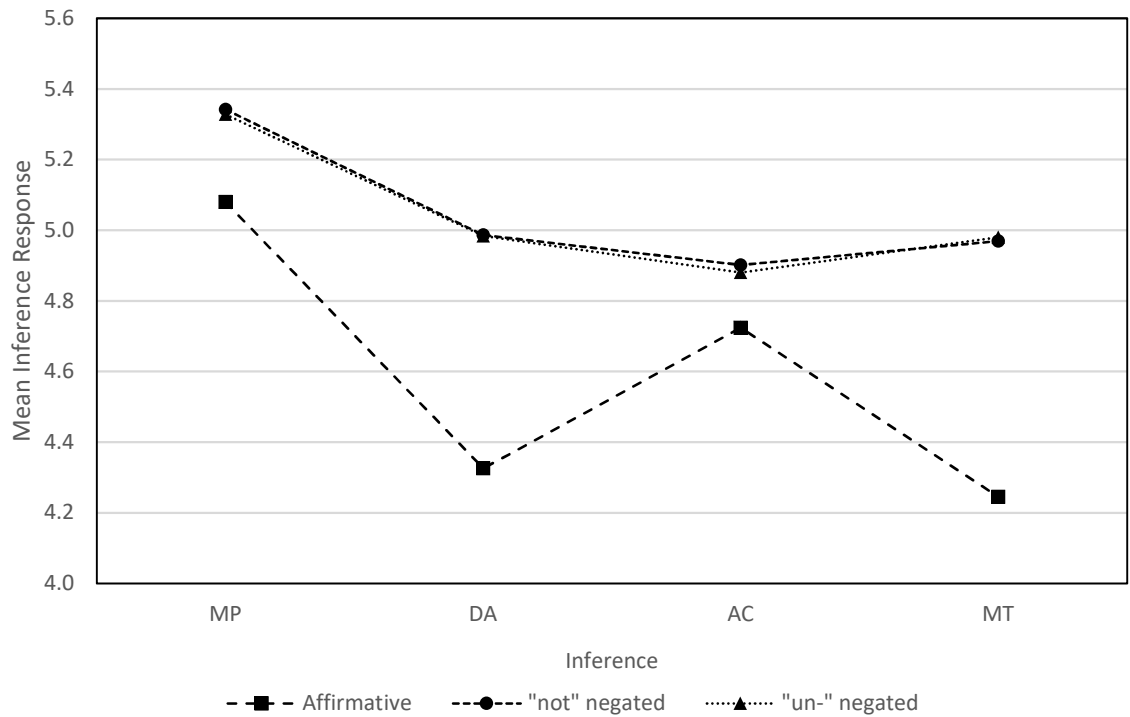


Figure 3.12 – Chart showing interaction between inference type and antecedent polarity on inference responses in experiment 4.

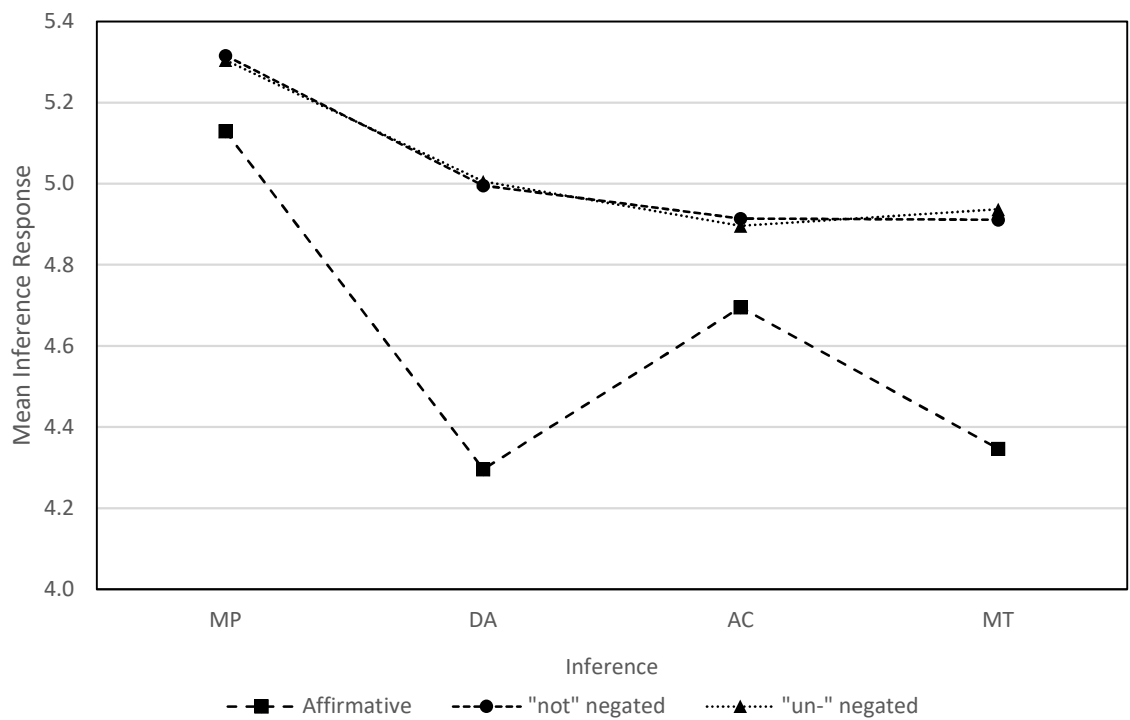
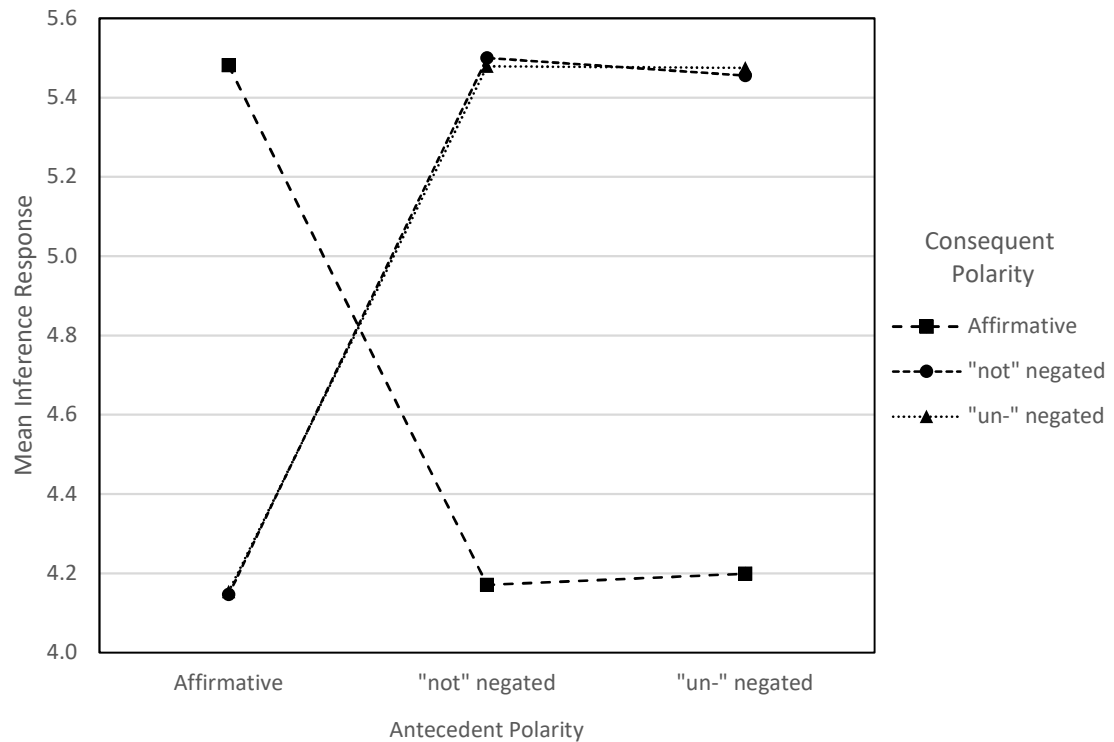


Figure 3.13 – Chart showing interaction between inference type and consequent polarity on inference responses in experiment 4.



*Figure 3.14* – Chart showing interaction between antecedent polarity and consequent polarity on inference responses in experiment 4.

Given the significant interactions, post hoc pairwise comparisons with Bonferroni adjustment for multiple comparisons were used to compare inference ratings for each condition for each of the three independent variables. For the inference conditions, MP was found to significantly vary from DA ( $p=.011$ ), AC ( $p=.001$ ) and MT ( $p=.005$ ). For antecedent polarity, significant differences were observed between affirmative and “not” negated ( $p<.001$ ) and affirmative and “un-” negated ( $p<.001$ ) conditions. Similarly, for consequent polarity, significant differences were observed between affirmative and “not” negated ( $p<.001$ ) and affirmative and “un-” negated ( $p<.001$ ) conditions.

### 3.6.3.2. Negative Conclusion Bias

The mean inference scores by inference type and conclusion polarity are shown in table 3.12.

*Table 3.12 – Mean (standard deviation) responses to inference questions in chapter 3, experiment 4*

Conclusion Polarity	Inference type			
	MP	DA	AC	MT
Affirmative	5.13 (2.03)	5.00 (1.91)	4.72 (1.92)	4.97 (1.89)
“not” negated	5.32 (1.96)	4.30 (1.91)	4.90 (1.90)	4.25 (1.90)
“un-” negated	5.30 (1.94)	-	4.88 (1.90)	-

Two one-way MANOVAs were conducted comparing inference response and plausibility rating scores for MP and AC inferences with conclusion polarity as the factor.

For MP, this showed a significant multivariate effect of conclusion polarity on both dependent variables ( $F(2,85) = 5.972$ ,  $p < .001$ ,  $\eta_p^2 = .219$ ) and univariate effects of conclusion polarity on inference response ( $F(2,176) = 7.083$ ,  $p = .001$ ,  $\eta_p^2 = .074$ ) and plausibility rating ( $F(2,176) = 9.984$ ,  $p < .001$ ,  $\eta_p^2 = .102$ ). Post hoc pairwise comparisons (with Bonferroni adjustment) were used to compare inference ratings for each conclusion polarity condition for MP. These showed significant differences in inference ratings between affirmative and “not” negated ( $p = .018$ ) and affirmative and “un-” negated conclusions ( $p = .017$ ). They also showed significant differences in plausibility ratings between affirmative and “not” negated ( $p = .004$ ) and affirmative and “un-” negated conclusions ( $p = .005$ ). No differences were found between the two different negation conditions for either dependent variable.

For AC, this showed a significant multivariate effect of conclusion polarity on both dependent variables ( $F(2,85) = 63.780$ ,  $p < .001$ ,  $\eta_p^2 = .750$ ) and univariate effects of conclusion polarity on inference response ( $F(2,176) = 5.667$ ,  $p = .004$ ,  $\eta_p^2 = .061$ ) and plausibility rating ( $F(2,176) = 132.702$ ,  $p < .001$ ,  $\eta_p^2 = .601$ ). Post hoc pairwise comparisons (with Bonferroni adjustment) were used to compare inference ratings for each conclusion polarity condition for MP. These showed significant differences in inference ratings between affirmative and “not” negated ( $p = .016$ ) and affirmative and “un-” negated conclusions ( $p = .029$ ). They also showed significant differences in plausibility ratings between affirmative and “not” negated ( $p < .001$ ), between affirmative and “un-” negated conclusions ( $p < .001$ ) and between “not” and “un-”

negated conclusions ( $p < .001$ ). No differences were found between the two different negation conditions for either dependent variable.

### 3.6.3.3. Role of Plausibility

The mean response to the probability rating tasks using the conditionals was 31.9 (standard deviation 26.3). The mean response to the probability rating tasks using the statements of antecedent propositions was 40.5 (standard deviation 25.5). The mean plausibility ratings given to each of the conditionals with each antecedent and consequent polarity are shown in table 3.13. The mean plausibility ratings given to each of the statements of antecedent propositions with each polarity are shown in table 3.14.

*Table 3.13 – Mean plausibility ratings of conditionals by antecedent polarity and consequent polarity*

Consequent Polarity	Antecedent Polarity		
	Affirmative	“not” negated	“un-” negated
Affirmative	61.5	21.4	21.8
“not” negated	21.8	34.6	34.9
“un-” negated	21.5	35.3	34.6

*Table 3.14 – Mean plausibility ratings of antecedent statements by polarity*

	Polarity		
	Affirmative	“not” negated	“un-” negated
Mean plausibility rating	63.9	29.2	28.4

As considered in the previous experiment, we can look at whether these ratings meet the following probability theory constraint:

$$P(\neg q \mid p) + P(q \mid p) = 1$$

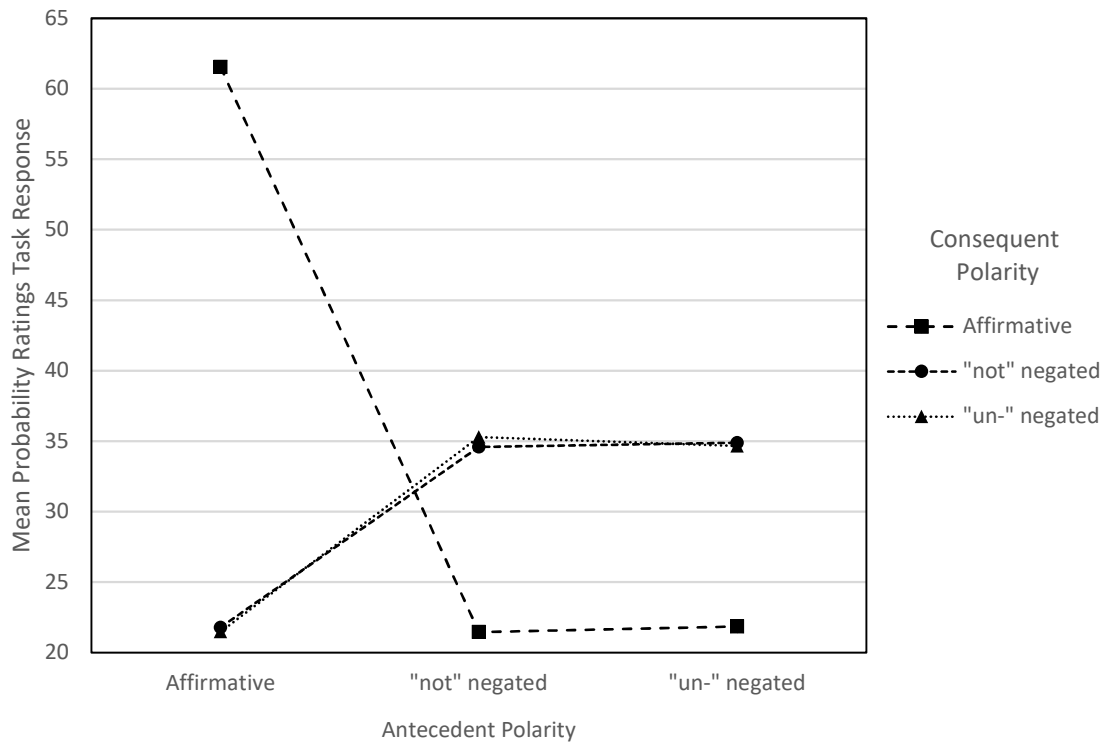
As before (§3.5.3.3) we can considering the sum of the plausibility ratings (divided by 100) for the affirmative consequent case and each of the cases with a negated consequent. On average these are .833 (for the “not” negated consequent case) and .830 (for the “un-” negated consequent case) where the antecedent is affirmative. Where antecedents are “not” negated, the equivalent values are .560 (for the “not” negated

consequent case) and .567 (for the “un-” negated consequent case). Where antecedents are “un-” negated, the equivalent values are .567 (for the “not” negated consequent case) and .564 (for the “un-” negated consequent case). All of these results are much less than 1, suggesting people’s average ratings provide perceived probabilities that are ‘superadditive’ (Macchi, Oshero & Kranz, 1999).

As set out above, a 4x3x3 within participants MANOVA was used to compare mean inference response scores and mean conditional plausibility scores as dependent variables from each participant across the factors inference type (MP, DA, AC, MT), antecedent polarity and consequent polarities (the latter two both have three levels: affirmative, “not” negated, “un-” negated). The significant multivariate effects are reported above along with the univariate effects on inference responses.

Univariate analysis for the effects of the independent variables on plausibility ratings found interactions between antecedent polarity and consequent polarity ( $F(4, 1056) = 170.085, p < .001, \eta_p^2 = .659$ ) shown in figure 3.15. It also found main effects of antecedent polarity ( $F(2, 1056) = 8.936, p < .001, \eta_p^2 = .092$ ) and consequent polarity ( $F(2, 1056) = 9.984, p < .001, \eta_p^2 = .102$ ).

Given the significant interactions, post hoc pairwise comparisons (with Bonferroni adjustment) were used to compare plausibility ratings between each level of the two polarity conditions. For antecedent polarity, significant differences were observed between affirmative and “not” negated ( $p = .007$ ) and affirmative and “un-” negated ( $p = .010$ ) conditions. Similarly, for consequent polarity, significant differences were observed between affirmative and “not” negated ( $p = .004$ ) and affirmative and “un-” negated ( $p = .005$ ) conditions.



*Figure 3.15* – Chart showing interaction effect between antecedent and consequent polarity on probability ratings of conditionals in experiment 4.

There was a significant positive correlation ( $r=.233$ ,  $p<.001$ ) between the perceived plausibility of the conditional statement (measured using the probability rating task score) and the level of agreement with the inference using that statement (measured using the response to the inference task). There was also a smaller significant negative correlation ( $r=-.088$ ,  $p<.001$ ) between the perceived plausibility of the antecedent statement (measured using the probability rating task score) and the level of agreement with the inference using that statement (measured using the response to the inference task).

#### 3.6.3.4. Other Possible Factors

The key additional factor that needs consideration for this experiment is the potential for scenario effects. There was only one participant who did not report having English as their first language so this the risk that this factor skewed results is very low compared to prior experiments. Participant factors are also addressed through using analyses appropriate to a fully within-participant design. The impact of moving from a categorical response to rating scale was considered as part of the previous experiment and found not to be a factor that substantially affected responses.

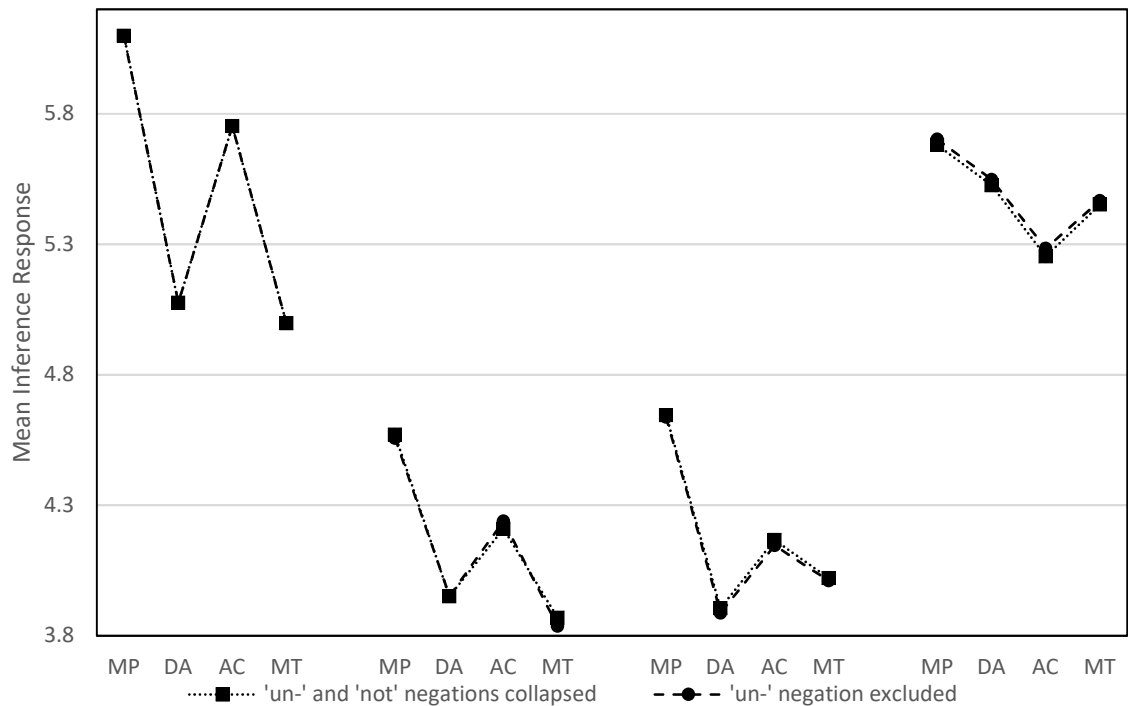


This experiment used four scenarios. A 4x4x3x3 MANOVA compared inference scores and plausibility ratings with the scenario, inference type, antecedent polarity and consequent polarity as factors. The same significant interactions and main effects (and no additional significant results) were found involving only the inference type, antecedent polarity and consequent polarity factors as the MANOVA described in §3.6.3.1 and §3.6.3.3 above (both the multivariate and univariate analyses). Additional multivariate effects on both dependent variables were found for interactions between scenario, antecedent polarity and consequent polarity ( $F(24,65) = 6.242, p < .001, \eta_p^2 = .697$ ), scenario and antecedent polarity ( $F(12,77) = 6.458, p < .001, \eta_p^2 = .502$ ) and scenario and consequent polarity ( $F(12,77) = 7.094, p < .001, \eta_p^2 = .525$ ). An additional multivariate main effect was found for scenario ( $F(6,83) = 8.673, p < .001, \eta_p^2 = .385$ ). Univariate analysis of the effects on inference responses found additional significant interactions between scenario, antecedent polarity and consequent polarity ( $F(12,3168) = 28.437, p < .001, \eta_p^2 = .244$ ), scenario and consequent polarity ( $F(6,3168) = 14.818, p < .001, \eta_p^2 = .144$ ) and scenario and antecedent polarity ( $F(6,3168) = 19.948, p < .001, \eta_p^2 = .185$ ). In addition a significant univariate main effect of scenario on inference responses was found ( $F(3,3168) = 21.493, p < .001, \eta_p^2 = .196$ ). Univariate analysis of the effects on inference responses found additional significant interactions between scenario, antecedent polarity and consequent polarity ( $F(12,3168) = 24.354, p < .001, \eta_p^2 = .217$ ), scenario and consequent polarity ( $F(6,3168) = 9.863, p < .001, \eta_p^2 = .101$ ) and scenario and antecedent polarity ( $F(6,3168) = 19.948, p = .018, \eta_p^2 = .028$ ). In addition a significant univariate main effect of scenario on plausibility ratings was found ( $F(3,3168) = 11.938, p < .001, \eta_p^2 = .119$ ). Post hoc pairwise comparisons (with Bonferroni adjustment) comparing inference ratings for each scenario condition found significant differences between scenario 1 and scenarios 2 ( $p < .001$ ) and 3 ( $p = .009$ ), scenario 2 and scenario 4 ( $p < .001$ ) and scenario 3 and scenario 4 ( $p < .001$ ). This appears to reflect scenarios 1 and 4 being closely correlated across conditions and 2 and 3 being closely correlated across conditions. All scenarios had a similar effect pattern on inferences, with higher inferences where the polarities of the antecedent and consequent were congruent (both affirmative or both negated – either negated condition). The difference between congruent and “non-”congruent polarity conditions is larger for scenarios 2 and 3 than for 1 and 4. Similar pairwise comparisons comparing plausibility ratings found significant differences between scenario 1 and scenario 2 ( $p = .002$ ), scenario 2 and scenario 4 ( $p < .001$ ) and scenario 3 and scenario 4 ( $p = .006$ ). The pattern

of the scenario effects is similar to that with inference responses (higher plausibility ratings where polarities are congruent). Again the difference looks larger with scenarios 2 and 3 (which have lower plausibility ratings where polarities are incongruent and higher ratings where polarities are congruent than 1 and 4). However, this difference is not significant in all cases.

#### **3.6.4. Discussion**

Like the previous experiment (see discussion in §3.5.4), the pattern of results differed from expectations. Like that experiment, patterns of inference did not completely align with the data from prior experiments where Evans' (1977) negations paradigm has been used (see figure 3.16). Like the previous experiment, inference levels are much lower where plausibility ratings are lowest (i.e. where the conditional has incongruent antecedent and consequent polarities). However, despite using the same scenarios, the pattern does not completely align with the previous experiment either. There does not appear to be the same suppression of MP which is now more endorsed than DA, AC and MT overall. Similarly, AC is now more endorsed overall than MT (although this normal pattern is still reversed where both antecedent and consequent use some form of negation).



*Figure 3.16* – Chart showing mean inference responses in experiment 4 by inference type and standard negation paradigm conditional formats.

The panels show (from left to right): (i) responses where conditional had affirmative antecedents and affirmative consequents; (ii) responses where conditionals had affirmative antecedents and negated consequents; (iii) responses where conditionals had negated antecedents and affirmative consequents; and (iv) responses where conditionals had negated antecedents and negated consequents. Different lines show inference responses with the results from conditionals using “un-” negated excluded and with responses to conditionals using the two different types of negation averaged.

Like the prior experiment, where differences between polarity conditions were observed, it generally showed that the effects of “un-” negated propositions was closer to the effects of “not” negated propositions than the effect of affirmative propositions.

Unlike the prior experiment, the analysis considering conclusion polarities did show higher levels of inference where conclusions were negated for MP and AC (though still not for DA and MT). MP and AC are the most interesting cases in this condition because they include conclusions in all three polarity conditions. For both of these inference types the order of the results aligned with the hypotheses (that inferences would be more readily endorsed when “not” negation was used in the conclusion and that where “un-” negation was used inference levels would fall between the levels seen for “not” negated and affirmative conclusions). However, the predicted significant difference in inference levels between conclusions using “not” and “un-” negation was

not found (although both negation conditions differed significantly from the affirmative condition).

The key difference between this experiment and the previous one was the extent to which participants will have had the opportunity to get used to the scenarios – they will have been exposed to each scenario 36 times in responding to inference questions and another 12 times for the plausibility rating questions. While the random presentation of the tasks will mitigate order effects, most responses were made after participants had seen multiple versions of the relevant scenario (rather than seeing each scenario for the first time as was the case in the previous experiment). Evidence that this led to different judgements about the conditionals includes the superadditive nature of the judgements about the likelihood of each conditional relationship (compared to responses in the prior experiment consistent with probability theory). As with the finding by Macchi et al. (1999) that people make more superadditive judgements where they have less knowledge of the events in question, greater exposure to the scenarios may have affected people's perceptions of them.

Greater familiarity with the form of the tasks may also have led to participants considering the different negation conditions alongside each other as they went through the experiment rather than in isolation. This may have led to them developing a strategy of responses which aligned the interpretation of the two different negations used, minimising any differences that might otherwise have been observed. This exposure to each scenario may also have allowed them to overcome the MP and AC suppression effects implied by the results to the previous experiment – greater familiarity with the form may have allowed these to be processed with less direct reference to the scenario in each case reducing the context effects seen previously. While MP inferences may have triggered more disablers when a scenario was considered for the first time, people may have taken the inference more for granted when considering it as part of a series of inferences. This will have led to the observed higher level of inference under the Fernbach and Erb (2013) CBN model.

Greater familiarity with the scenarios (and a lack of participant difference factors in a within-participants design) may also have allowed scenario effects to be observed more clearly. In this experiment, we saw that the two scenarios with, arguably, better related

propositions (scenarios 2 – which related ‘friendly’ and ‘social’ – and 3 – which related ‘grateful’ and ‘happy’) showing a more extreme range of plausibility between congruent and incongruent polarity conditions. This greater difference in plausibility related to greater differences in inference response levels.

As before, and as expected, plausibility ratings of the conditionals, does appear to be a key factor in people’s willingness to endorse inferences using them. While the pattern of results did not match predictions, the presence of this factor means that these results continue to provide strong support for theories based on probabilistic (e.g. Oaksford et al., 2000) and CBN (e.g. Fernbach & Erb, 2013) models. That differences in probability perceptions between this experiment and the previous one appeared to affect inference judgements is also consistent with these theories. These effects are harder to explain with reference to mental models (e.g. Johnson-Laird & Byrne, 2002) where the mental models considered should have been identical between the two experiments.

This experiment did collect a large amount of data through a large within-participants design. While the results are not aligned with the predicted effects of different types of negation, the data – and the differences between that obtained in the prior experiment – continue to provide broad support for the role of probabilistic and CBN models in reasoning.

### **3.7. General Discussion**

The four experiments discussed in this chapter aimed to explore how the use of different types of negation in conditional inference tasks might affect reasoning biases. The experiments did not identify any robust effects that emerge when different types of negation are used. They may not have been sufficiently sensitive to identify what are likely to be small effects. However, the experiments all provided substantial evidence for the role of plausibility in people’s inference judgements.

Experiments 1, 3 and 4 showed higher rates of inference endorsement when conditionals rated more plausible were used. This is the same pattern as observed in previous studies which have considered the role of plausibility in conditional reasoning (e.g. Quinn & Markovits, 1998). In these experiments, the plausibility of the conditional was not actively manipulated (although the scenarios used in experiments 3 and 4 were

designed to create a natural profile of plausibility). There may be other factors that affect both the variance in plausibility between the conditionals and inference rates – such as the interaction effect of antecedent and consequent polarity factors on both plausibility ratings and inference responses. However, these associated factors may also suggest a mechanism by which the plausibility of conditionals (which is likely to be different for most conditional when the form ‘if  $x$ , then  $y$ ’ is changed to ‘if  $x$ , then not  $y$ ’) is partly responsible for negation effects seen elsewhere. So, people may consider ‘if  $x$ , then  $y$ ’ a plausible relationship (either because of what they know about  $x$  and  $y$ , their salience to each other or because of the information provided in context). The relationship ‘if  $x$ , then not  $y$ ’ will then be considered less plausible because it represents, in some respects, the opposite of a relationship they consider plausible. Based on the differences we have seen when plausibility is higher, inferences are more likely to be drawn when the more plausible ‘if  $x$ , then  $y$ ’ is used than the less plausible ‘if  $x$ , then not  $y$ ’. This may lead to the observation of effects apparently related to the insertion of negation which is actually caused by plausibility and the effect the insertion of negation has on the plausibility of the conditional. These experiments do not allow us to test this proposition, but their results would appear to be consistent with it.

Another finding that emphasises the role of plausibility in inference tasks is the odd results in experiments 2. Here conditionals with low variation in plausibility ratings were used and these may have had ambiguous plausibility. These scenarios led to the odd result that no difference between DA, AC and MT inference levels were observed. This may be because the typical inference type effects were undermined by participants’ confusion at dealing with scenarios with ambiguous plausibility. This, again, potentially emphasises the importance of conditional plausibility as a factor in inference responses (even compared to well-established inference type effects). A similar suppression of inference effects was seen in experiment 3, but the fact that MP inference rates were also much lower than normal (and that this effect was not seen in experiment 4) suggests that another factor may have been responsible – possibly related to how people consider disablers when provided with these conditionals for the first time (rather than once they are used to the scenarios).

The results of the experiments in this chapter did not demonstrate significant effects of different types of negation on reasoning with conditional inference tasks. However, the

experiments do not rule out the possibility that small effects exist. It appears that, in the tests conducted, other factors were more significant in influencing reasoning responses – a key factor appears to be the plausibility of conditionals. The impact of this factor provides substantial support for probabilistic and CBN models of reasoning (e.g. Oaksford et al., 2000; Fernbach & Erb, 2013).

To continue consideration of the effects of different types of negation in reasoning tasks, it might be interesting to see whether a larger study in which people only saw each scenario once replicated the results of experiment 3 (and the differences with experiment 4). Different plausibility manipulations could be used (for example adjustments to the likelihood profiles suggested by scenarios) which might reduce plausibility effects and make negation effects easier to observe. However, it is clear from the experiments conducted that plausibility has a very substantial effect on results and it is not easy to mitigate this. Therefore, it seems sensible to move onto to consider another aspect of the role of negation in reasoning.

## **4. Contrast Classes in Conditional Inference**

Previous chapters have looked at whether different understandings of negation can help us understand and interpret the reasoning biases that occur when using negation. In the current chapter, I will consider a further understanding of negation based on the philosophical account of negation as implying ‘otherness’ (Horn, 1989). This philosophical account informs the psychological account of ‘contrast classes’ (Oaksford, 2002; Oaksford & Stenning, 1992). The contrast class account has been used to underpin the optimal data selection account of the effects of negation in Wason’s (1968) selection task (Oaksford, 2002; Oaksford & Stenning, 1992; Oaksford & Chater, 1994, 2003a, 2007; Oaksford and Moussakowski, 2004; Singmann, Klauer & Beller, 2016). In the present chapter, I consider the influence of contrast classes in people’s responses to conditional inference tasks and how that might help us understand how people reason with negations.

### **4.1. The Contrast Class Account of Negation**

The philosophical depiction of negation as otherness – that a negated statement is seen as a positive reference to something other than the negated proposition – can be traced back to Plato (Horn, 1989). The otherness approach does not appear to be able to account for all uses of negation (for example, an instruction using negation like ‘do not kill’ has a different implication to any attempt to assign an entirely positive counterpart, e.g. ‘let all live’). However, the approach may provide another way to understand and process negations under certain circumstances – where positive alternatives to negated statements are readily available.

As discussed earlier (§1.4.3), people form cognitive groupings, reference sets, into which they categorise items. People’s ability to form and use categories around concepts was demonstrated by Rosch (1973) in a pair of experiments which trained monolingual participants whose language did not describe differences in colour and geometric forms to identify sets of colours and geometric forms. She found that participants more quickly learned to identify sets with a ‘presumed natural prototype’. For example, for colours, sets with members that were closely related to a colour (in terms of hue and brightness) which had been identified as normally relating to a basic colour term in prior cross-language studies (e.g. ‘blue’ in English) were easier to learn



than categories centred around colours that did not match a normal basic colour term (e.g. 'red and brown' in English). Once learned, Rosch (1973) demonstrated that almost all participants could extend the set to apply to a previously unseen stimulus which was consistent with the set.

Therefore, people normally have ready cognitive access to a set of items which might be appropriate and relevant to a given proposition. For example, when a proposition considers modes of travel between the UK and France, a group of items likely to include 'aeroplane', 'ferry' and 'train' will be considered more readily than a wider group of modes of travel (which could also include 'walking' or 'bicycle'). This group of items provides a reference class that people can use when considering a proposition. So, when a person considers a journey between London and Paris, a group of options will be considered that may well include 'train' but is unlikely to include 'bicycle'. Similarly, on consideration of such a journey, an individual will narrow the reference set considered to relevant modes of transport and exclude from consideration items that are not relevant (like 'cat' which may be considered within categories of 'animals' or 'pets' but not a category which requires modes of travel).

Rosch and Mervis (1975) conducted a series of experiments that investigated the structure of cognitively formed reference sets. They found that items considered most prototypical of a class were those that had the most attributes in common with other members of the class and the least attributes in common with members of other classes. As part of these experiments, they looked at how participants could generate contrast sets of items that are in the same category as an item (they share relevant attributes) but are not the item. This demonstrated that, when negating an item, people can use reference sets of cognitively related items to form relevant contrast sets. For example, if you asked your friend whether they were taking the train on their trip from London to Paris and they clarified that they were not, then you would probably consider options like aeroplane and ferry (which form part of a reference set of ways to get to France) as alternative possibilities. Note that such contrast sets are cognitive constructs, not logical ones as they exclude items not likely to be relevant to the proposition in hand. The logical contrast set of items which are not taking a train would include riding a cat and swimming. The former is excluded from most people's cognitive contrast set as not feasible and the latter would normally be excluded as highly unlikely (except in certain

contexts – such as a fit friend planning a highly exhausting trip to France to raise money for charity!)

Cognitive contrast sets that include items similar to (but which are not) a given item are called contrast classes of the given item (Oaksford & Stenning, 1992). The probabilistic approach to reasoning has considered how people might use contrast classes in reasoning (Oaksford, 2002).

#### **4.1.1. Contrast Classes in Reasoning**

Oaksford and Stenning (1992) sought to explain matching bias in Evans' (1972a) construction task and Wason's (1968) selection task by considering how negated propositions are understood as identifying a contrast class rather than simple denial. They contrasted this 'processing negations' account with the heuristic-analytic account (e.g. Evans, 1984) which suggested that matching was caused by a rejection of irrelevant (non-matching) information at an initial heuristic stage before the problem was considered logically in a subsequent analytic stage. Their experiments predicted that matching bias would be suppressed in conditions which eased contrast class construction (by providing a context in which a contrast class can easily be derived rather than an abstract form where the nature of the contrast class is ambiguous). The heuristic-analytic approach would not anticipate any suppression of matching effect in such conditions. Oaksford and Stenning (1992) observed effects which supported their processing negation account and not the heuristic-analytic account, providing an initial indication that contrast classes may play a role in people's reasoning.

Oaksford and Chater (1994) set out a more developed probabilistic approach which predicts responses to Wason's (1968) selection task. This approach proposed that people's responses are based on seeing the task as one of optimal data selection (selecting the answer that would most likely provide the most information about whether the given rule was true or false) rather than the logic of falsification. They developed a model which predicted the cards chosen in different circumstances based on the likelihood that the card would provide useful information about whether the rule was true or false. The model assumed that people would interpret negated items with reference to the item's contrast class. They anticipated that people would ascribe a higher probability to the contrast class of an item (that the item was not the case) than to

the single item itself because there is often more than one item in a contrast class. Also, they set the probability that an item is true or that a member of its contrast class is true at less than one. This is because the plausible contrast class is not the full complement of possibilities to the negated item. Therefore, there remains a small possibility that it is not the case that either the item is true or a member of the plausibly generated contrast class is true.

Oaksford and Chater (1994) found that their model was a good match for data derived from prior experiments. They, therefore, concluded that responses to the selection task were likely a rational response based on an adaptive approach to hypothesis testing which would be more appropriate in everyday life rather than a systematic bias against classical logic. The good fits obtained for a model which assumed contrast class construction support the hypothesis that people interpret negation in reasoning with reference to appropriate contrast classes.

Further indirect evidence for people's use of contrast classes to interpret negation in reasoning problems was observed by Oaksford, Chater and Larkin (2000). They proposed a probabilistic model which predicted responses to conditional inference tasks. As with their model for Wason's (1968) selection task (Oaksford & Chater, 1994), Oaksford et al. (2000) suggested that people normally assume that a negated proposition would have higher probability than the same proposition without negation. This is because the contrast class associated with the proposition would contain more possibilities (and therefore be more likely) than the proposition itself. The model considered the conditional probabilities implied by conditional statements and predicted that inferences would be more highly endorsed where their likelihood was high (rather than necessarily where they are logically correct). This model provided a good fit for a meta-analysis of prior experimental data. The theory proposed that negation effects were due to the interpretation of the negated proposition as having a higher probability than the affirmative proposition (because of the greater number of possibilities in the contrast class) was also supported by three experiments. The experiments tested high probability affirmative statements in place of negated propositions. They found that these high probability propositions led to effects similar to those previously seen when using negations in abstract tasks (including one that used scenarios similar to many prior tasks) and a task with naturalistic scenarios.

Initial studies (Oaksford & Stenning, 1992; Oaksford & Chater, 1994; Oaksford et al., 2000) proposing the relevance of contrast classes to reasoning had used the concept to inform probabilistic models. Meta-analyses, fitting the model to prior data and new experiments suggested that these models had some predictive power, supporting the case that contrast classes are relevant to people's understanding of negation in reasoning tasks. However, these studies had not shown direct evidence for contrast class generation when people approached reasoning problems.

The heuristic-analytic theory's (e.g. Evans, 1984) explanation of bias in conditional reasoning continues to provide an alternative to probabilistic approaches which assume contrast classes are relevant to interpretations of negations. This approach suggests that people did consider tasks logically and responded accordingly following an analytical stage and that a prior heuristic stage which selected only information deemed relevant for consideration (to reduce the requirements of the more complex analytic phase) causes illogical biases. Negation is therefore not considered at the heuristic stage (which simply looks at whether a proposition appears relevant to the conditional being considered – e.g. the content of the proposition matches a term in the conditional. At the analytical stage, the problem is considered with negation representing simple denial (rather than representing something other than the subject of the negation as required by the contrast class interpretation). Evans (1998) highlighted a series of experiments by Evans, Clibbens and Rood (1996) that demonstrated that the use of explicit negations (rather than implicit negations) in Wason's (1968) selection task appeared to remove matching bias. Evans (1998) suggested that the heuristic-analytic account with an initial step to narrow the information considered to that relevant (Evans, 1984) and processing negations account using contrast classes (Oaksford & Stenning, 1992) were both more likely to account for these findings than other theories. For example, Evans (1998) considered that mental models accounts (e.g. Johnson-Laird & Byrne, 1993) did not easily predict the observed matching of negated antecedent terms in the selection task – a matching of terms which would not be included in the initial mental model.

Some experiments have compared the differing predictions that the dual process approach (e.g. Evans, 1984) and the contrast class approach (e.g. Oaksford and Chater, 1994) make for responses to the selection task. Yama (2001) conducted a series of

experiments which provided participants with frequency data about alternative options and asked them to complete a version of the selection task using materials related to this information. Essentially, Yama (2001) predicted that if people took a probabilistic approach to the selection task, differences in the frequency data presented would have a greater impact on responses than the matching bias anticipated by the heuristic-analytic approach. These experiments suggested that matching bias was not suppressed by the provision of frequency information which would reduce the probabilistic motivation for matching bias under Oaksford and Chater's (1994) optimal data selection approach. This potentially undermined the case that people interpreted negations in reasoning with reference to contrast classes. However, Oaksford (2002) argued that Yama's (2001) results supported a larger number of the detailed predictions made by the contrast class account than of those made by the dual-process account. Oaksford and Moussakowski (2004) followed this with a series of three experiments that provided the frequency information in a manner more likely to engage participants. After first replicating Yama's (2001) key experiment with a group of western participants, they conducted two experiments which provided frequency information in a more naturalistic way. People do not normally derive their beliefs about the likelihood of events from tables of information but through sequential sampling – observing a single case, one at a time, many times (Oaksford & Wakefield, 2003). Oaksford and Moussakowski (2004) therefore provided information about frequencies by presenting examples of occurrences one at a time. Unlike Yama (2001), they did find that the frequency information had the effect predicted by the optimal data selection approach. This suggests that where the presentation of information reflects the manner that they naturally absorb it, people's reasoning judgements to reflect probabilistic consideration.

Several studies have provided evidence which undermines the contrast class account, including Prado and Noveck (2006) and Stahl, Klauer and Erdfielder (2008).

Prado and Noveck (2006) sought to compare the relevance of 'narrow' and 'search for alternatives' interpretations of negation to matching bias. They described the 'narrow' view of negation as logical denial and the 'search for alternatives' was described in line with the contrast class account (Oaksford & Stenning, 1992). They used an evaluation task (in which participants consider whether a given stimuli are consistent with a conditional rule) and a falsification task (where participants consider whether a given

stimuli would falsify the given conditional). Focusing on conditionals with affirmative antecedents and negative consequents (the form *if p then not-q*), they found that participants were both more successful (significantly so) and quicker (although not significantly quicker) at correctly rejecting stimuli which matched the terms of the rule (i.e. consistent with *p* and *q*) than at confirming the validity of stimuli consistent with the rule (i.e. showing *p* and *not-q*). Similarly, in the falsification task, they found that participants were significantly quicker to recognise stimuli that falsified the rule (showing *p* and *q*) than to correctly reject stimuli that would not (showing *p* and *not-q*). There was no significant difference in accuracy of responses between these two conditions in the falsification task. Prado and Noveck (2006) argued that this was consistent with their ‘narrow’ view of negation and Evans (1998) proposed matching heuristic. They suggested that a ‘search for alternatives’ approach to negation would predict greater accuracy and quicker responses where the stimuli included an example of *not-q* rather than *q* itself. They found similar support for the hypothesis based on the ‘narrow’ view by showing that people would more quickly and accurately respond where stimuli ‘matched’ the terms of the rule (i.e. showed *p* and *q* regardless of the negation in the conditional) when comparing trials with stimuli that falsified that affirmative antecedent and affirmative consequent rule (*if p then q*) to trials with stimuli that confirmed the affirmative antecedent and negative consequent version of the rule (*if p then not-q*).

Stahl, Klauer and Erdfielder (2008) compared the function of implicit and explicit negations when used in Wason’s (1968) selection task with propositions involving numbers and letters. They found that the use of explicit negation reduced the matching bias observed but did not remove it as had been observed in prior studies (e.g. Evans, 1996). This finding undermines basic heuristic accounts (e.g. Evans, 1984) which rely on a ‘relevance’ heuristic prioritising consideration of items named in the conditional given to account for matching bias (as, when using explicit negations, the inference does not state the item matched). The authors also tried and failed to model their data using Oaksford and Chater’s (2003) optimal data selection model which relies on contrast class understandings of negation. These results call into question these approaches. Stahl et al. (2008) were able to model their data using an updated version of Evans (2006) heuristic-analytic theory that allowed the use of negations in the task to affect the task’s resolution strategy.

Prado and Noveck (2006) and Stahl, Klauer and Erdfielder (2008) provide evidence which may suggest that contrast classes are less relevant to reasoning with negations. However, Prado and Noveck's (2006) expectation that accessing contrast classes would create a quicker response when a participant looked at a member of the contrast class rather than the named item itself may not be correct. Reference to a contrast class may be used to support processing of reasoning and may lead to contrast class members being processed and considered more quickly than irrelevant items. However, that does not mean contrast class members can be processed more quickly than the item named which is also likely to be part of the individual's processing. Therefore, it may not be reasonable to dismiss the contrast class account on the basis of the experiment conducted by Prado and Noveck (2006). Stahl et al. (2008) also do not rule out the relevance of an optimal data selection model using contrast classes to explain matching bias in the selection task. It may be that the model they used needs to be better adapted to address the requirements of their experiment.

More recently, Gale and Ball (2012) investigated the role of contrast classes in Wason's (1960) rule discovery task. This task gives participants a 'triple' consisting of three numbers (traditionally '2-4-6') which participants are told is consistent with a rule that the experimenter is using. Participants are invited to suggest further triples which the experimenter tells them are either consistent or inconsistent with the rule. The participants' task is to identify the rule and Wason (1960) suggested that the narrow rules many proposed (e.g. 'numbers increasing by two') were an example of 'confirmation bias' as many of these participants had only suggested triples consistent with that hypothesis. Those participants that tried triples inconsistent with their hypothesised rule were more likely to uncover the true rule, 'increasing numbers'. Subsequent researchers (e.g. Sperber, Cara & Girotto, 1995) have argued that the exemplar leads participants to a narrower than necessary hypothesis, an example of 'relevance theory'. Success rates did increase where participants responded to a dual goal task (e.g. Gale & Ball, 2006). This told them that there are two rules and the example only represents one of them (all those that don't meet the 'increasing numbers' rule are said to conform to the alternative rule). Gale and Ball (2012) conducted three experiments where participants completed dual goal tasks. In their first two experiments participants were given '2-4-6' as the example of the 'increasing numbers' rule as usual

as well as an example of the other rule. In the first experiment, this example was from the contrast set of the initial rule and was either ‘useful’ (‘6-4-2’) or ‘nonuseful’ (‘4-4-4’). In their second experiment a third condition was introduced in which a ‘multiple-dimensions’ example for the other rule was provided (‘9-8-1’). They noted that the ‘useful’ example was the most likely member of a contrast class for the other example given (as defined by Oaksford, 2002, as the likely alternatives rather than the set of all alternatives). They found that around three-quarters of participants given the ‘useful’ exemplar were able to provide the ‘increasing numbers’ rule in the given amount of time (compared to about a fifth when using the ‘nonuseful’ rule and about a third with the ‘multiple dimensions’ rule). This suggests that prompts involving likely contrast class members are particularly facilitative at helping people solve this task. In their third experiment, Gale and Ball (2012) gave all participants the standard exemplar for the ‘increasing numbers’ rule and the ‘useful’ exemplar of the other rule. They asked participants what they thought the first rule was immediately after being provided with the task and again after they had suggested and had feedback on a set of alternative triples. Few participants guessed the rule initially, but most got it after the additional feedback on their triples. This suggests that the facilitation effect of providing a contrast class member was on the participants’ use of testing other triples to find the rules rather than on the initial guess. Overall Gale and Ball (2012) show how contrast class use can facilitate reasoning in the rule discovering task – demonstrating the wider application of contrast classes to reasoning problems.

#### ***4.1.2. Implicit Negations Using Contrast Class Members***

I discussed the use of implicit negation earlier (see §2.2.2), which Tottie (1982) described as a statement that negated a proposition without explicitly referencing the negated proposition.

Evans and Handley (1999) conducted three experiments using implicit negations in inference tasks. They used abstract materials referring to cards with numbers and letters on them (as are typically used in the selection task). In the first two experiments, they presented a conditional relationship and then a series of premises and conclusions (representing MP, DA, AC and MT) and asked whether the conclusion followed the premise in each case. In the third experiment, participants were just given the conditional and premise and asked to generate their own conclusion. The premises



given either used explicit negation (e.g. ‘A card which does not have a P’ where the conditional referred to cards with a ‘P’) or implicit negation (e.g. ‘A card which has a B’ given the same rule). This task showed that people were less likely to endorse (or generate a conclusion for) a conditional inference where the premises given were implicitly rather than explicitly negated.

Schroyens, Verschueren, Schaeken and d’Ydewalle (2000) conducted an experiment similar to Evans and Handley (1999), but they provided information about the size of the contrast class of the propositions in the conditional (i.e. what alternative values it could take). They found that the presence of a larger contrast class decreased the implicit negation effect which had been observed by Evans and Handley (1999). Schroyens et al. (2000) pointed out that this was consistent with Oaksford and Stenning’s (1992) processing negations account which suggested that people will reason with reference to the contrast class. Under the processing negations account (Oaksford and Stenning, 1992), explicit negation should readily trigger reference to the relevant contrast class. This will normally lead to the negation of a low probability affirmative proposition creating a high probability proposition based on the large number of alternatives that could be the case. Implicit negation may not trigger contrast class construction and therefore a proposition using implicit negation is likely to be seen as low probability – inhibiting inferences using it when it is the premise. Therefore, when the contrast class was manipulated to be large, the effects disappeared. In their second experiment, Schroyens et al. (2000) tested this by using implicit denials that either facilitated or did not facilitate reference to contrast class members. They again observed that implicit negation effects reduced when facilitating contrast class consideration.

Oaksford, Chater and Larkin (2000) introduced the conditional probability model for conditional inference (see §1.3.3 for a description of this model). The model is dependent on a contrast class approach to processing negations. While this model proved a good fit for most data, Oaksford et al. (2000) themselves found weaker support for the model in their second experiment which used implicit negations.

Schroyens and Schaeken (2003) provided a critique of Oaksford et al.’s (2000) model. They suggested that the conditional probability model should predict equivalent levels of endorsement based on  $P(p \mid r)$  for the following AC inferences:

*If  $p$ , then not  $q$ ;  $r \therefore p$*

*If  $p$ , then  $r$ ;  $r \therefore p$*

Schroyens and Schaeken (2003) pointed out that this did not align with the results seen in Evans and Handley (1999) and Schroyens et al. (2000). These studies saw implicit negation bias suggesting that these two inferences were not treated equivalently (with the implicit negation used in the first inference suppressing endorsements compared to the second, affirmative, version).

Oaksford and Chater (2003) point out that although these conditional inferences are superficially similar, these are not the conditionals that participants get to see in these experiments, which retain the same lexical content in antecedent and consequent positions. They point out that one of the examples of these inferences that Schroyens and Schaeken (2003) consider ('If you are in Paris, then you are in France. You are in France. Therefore, you are in Paris') has probability greater than zero. However, the corresponding statement that uses the same lexical content in antecedent and consequent ('If you are in Paris, then you are not in France. You are in England. Therefore, you are in Paris') does have a probability of zero. Therefore, the conditional probability is only the same where the lexical materials are allowed to vary between antecedent and consequent ('If you are in Paris, then you are not in England. You are in France. Therefore, you are in Paris') which is not the case in these experiments.

Oaksford and Chater (2007) further considered how the effects of implicit negation using members of a contrast class might still reflect rational responses, dependent on conditional probabilities. Such considerations may not have been fully incorporated into the basic model of Oaksford et al. (2000) which generalised treatment of contrast classes. However, the basic theory that inference is dependent on conditional probability could incorporate these considerations. They considered how MP would be considered differently in two superficially similar cases. They proposed the contingencies in table 4.1. Panel A ('Explicit negations') considers a dependency between propositions  $p$  and  $a$ . Panel B ('Implicit negations') represents the same dependency but splits out the *not- $p$*  cases as a contrast class (consisting of  $q$  and  $r$ ) and the *not- $a$*  cases into a contrast class (consisting of  $b$  and  $c$ ). Assuming those contrast classes are complete, then the

contingencies for *b* and *c* in panel B can be collapsed into the contingencies for *not-a* in panel A and, similarly, you can collapse *q* and *r* into *not-p*.

*Table 4.1 – Oaksford and Chater’s (2007) illustrative examples of contingency tables for a conditional rule where, e.g.  $\neg a = \{b\ c\}$*

Explicit negations			Implicit Negations		
	<i>a</i>	$\neg a$		<i>a</i>	<i>b</i> <i>c</i>
<i>p</i>	.3	.1	<i>p</i>	.3	.05    .05
$\neg p$	.1	.5	<i>q</i>	.1	.025   .025
			<i>r</i>	0	.225   .225

Source: Oaksford and Chater (2007), Table 5.2

We can assess the following inferences using explicit negations as set out in these contingency tables:

$$\text{If } \neg p, \text{ then } \neg a; \neg p \therefore a$$

$$\text{If } \neg p, \text{ then } \neg a; \neg p \therefore \neg a$$

According to the conditional probability model, endorsement of these inferences will be based on  $P(a \mid \neg p)$  and  $P(\neg a \mid \neg p)$  respectively. Based on either of the contingency tables in panel A or panel B, this provides probabilities of .167 and .833 respectively. The model would, therefore, suggest that the first inference (which is not logically correct) would see lower levels of endorsement than the second, MP, inference. Given a choice between these inferences, you would expect them to pick the latter, MP, inference in preference to the first.

However, we may see different conditional probabilities if we use members of a contrast class to provide implicit negation, as in the following two inferences:

$$\text{If } \neg p, \text{ then } \neg a; q \therefore a$$

$$\text{If } \neg p, \text{ then } \neg a; q \therefore \neg a$$

Although superficially similar to the inferences above, the use of a contrast class member (which is a member of the  $\neg p$  set) invokes different conditional probabilities:  $P(a \mid q)$  and  $P(\neg a \mid q)$  respectively. These have to be evaluated using the contingencies in panel B and now the probability of the first (logically incorrect) inference is .667, higher than the probability of the MP inference which is .333. In this case, the

conditional probability model would predict the choice of the first inference over the latter, MP, inference.

With this example, Oaksford and Chater (2007) provide an account of how the conditional probability model could account for implicit negation effects, such as those seen by Evans and Handley (1999) and Schroyens et al. (2000). This account would predict that you can create implicit negation effects by providing people with a reference contrast class and manipulating the frequencies involved. However, this prediction has not yet been directly tested.

The purpose of the current chapter is therefore to test Oaksford and Chater's (2007) intuitive understanding of reasoning with implicit negations using contrast class members. If manipulation of contrast classes in the manner proposed above does lead to suppression of MP inferences, then this will provide support for the notion (central to Oaksford & Chater's, 2000, model) that inference levels depend on conditional probability.

#### ***4.1.3. Implicit negation effects and other theories of reasoning***

As discussed above, a finding that manipulating frequency information related to contrast class members suppressed MP inferences would support the underlying theory of Oaksford and Chater's (2000) conditional probability model for conditional inference.

If the use of implicit negation can affect inferences when they imply a different conditional probability for the inference, then it is not clear how mental models theory (e.g. Johnson-Laird & Byrne, 2002) can account for this result. Given the rule 'If not  $p$ , then not  $q$ ' we would expect the generation of the following initial mental model:

$$\begin{array}{l} [\neg p] \quad \neg q \\ \dots \end{array}$$

This would allow the MP inference to be drawn quickly when using explicit negation. Superficially, the need to convert an implicitly negated premise into the 'not  $p$ ' term in the mental model may suppress inferences when using implicit negation. However, this

suppression effect would be the same regardless of the frequency data provided (so you should not expect different levels of suppression effect based on frequency data).

The fully fleshed out mental model for ‘If not  $p$ , then not  $q$ ’ is as follows:

$[\neg p]$	$\neg q$
$p$	$[q]$
$p$	$\neg q$

Based on this, a similar concern exists for the other forms of inference that we could test (DA, AC and MT).

Khemlani, Orenes and Johnson-Laird (2012) provide a more recent account of negation based on mental models theory. They note that negation of a proposition often leads to a wider range of possibilities and that modelling all possibilities could place a burden on mental capacity. For example, they note that modelling the possibilities for the statement ‘The students are not male adults’ could represent three narrow scope possibilities, modelled as:

<i>Not male</i>	<i>Adults</i>	<i>(i.e. the students are women)</i>
<i>Male</i>	<i>Not adults</i>	<i>(i.e. the students are boys)</i>
<i>Not male</i>	<i>Not adults</i>	<i>(i.e. the students are girls)</i>

Khemlani et al. (2012) therefore propose that a general effect of context on model generation will make it easier (i.e. quicker) to understand negation when people are already considering models of the corresponding affirmative assertion. They propose that people will generally take a narrow scope interpretation of negation (i.e. only the possibility set out in the model above that is best supported by context) to simplify model generation. Because the use of implicit negation does not refer to the corresponding affirmative assertion, they predict that it is, therefore, harder to understand (i.e. slower) than explicit negation using “not”. Khemlani et al. (2012) therefore propose a mechanism where consideration of the mental model is made more complicated by the requirement to translate an implicit negation into an explicit negation as used by the model. However, this effect would appear to be consistent, regardless of any frequency manipulations. Therefore, while Khemlani et al. (2012) find

supporting evidence for their predictions, their account cannot predict the effects anticipated by Oaksford and Chater (2007), set out above.

A heuristic-analytic theory (e.g. Evans, 1984) suffers from a similar constraint. Evans (2006) provides an extended heuristic-analytic account which suggests that a heuristic system unconsciously generates plausible models and either refers them to an analytic system for full consideration or uses them to inform immediate responses. Systematic biases, like matching, are attributed to this system, which provides quick responses based on requirements like relevance where required. The relevance requirement in the heuristic system could suggest that inferences are more likely to be rejected because the presence of an implicitly negative term is not obviously relevant to the conditional under consideration. However, such an effect would be expected to be flat and apply to all uses of implicit negation – not just the uses where the underlying frequency information also supported suppression. Evans' (2006) analytical system provides an explicit reasoning process for evaluating models and using them to either respond or refer back to the heuristic system for a new model. Evans (2006) argues that this analytical system can adopt a range of strategies depending on the nature of the problem under consideration. The ability of such a heuristic-analytic approach to account for the predicted effects, therefore, depends on the nature of the analytic system that you anticipate being employed (such as a probabilistic reasoning system like that proposed by Oaksford & Chater, 2000, or one based on Johnson-Laird & Byrne's, 2002, mental models).

## **4.2. Experiment 1**

This first experiment is intended to provide a basic test of the predictions set out in Oaksford and Chater (2007). Specifically, participants see frequency information based on Oaksford and Chater's (2007) contingency table (table 4.1 above). This should create a situation where endorsement of MP inferences depends on whether explicit or implicit negation – referring to a contrast class member – is used.

### ***4.2.1. Approach to providing contextual information***

Previous experiments have sought to provide context information intended to inform participants' conditional reasoning on a single page through narrative statements or data tables (e.g. Oberauer, Wilhelm & Rosas Diaz, 1999; Yama, 2001; Oaksford et al.,

2000). However, these approaches do not reflect the way in which people normally learn about the relative frequencies of occurrences in the world. Typically, people will experience different instances that relate to a subject under consideration over time. By seeing one instance, then another and another, people slowly build up an impression of the proportion of instances of each type.

Pollard and Evans (1983) conducted a selection task experiment which showed participants both sides of each card in a deck of 22, one at a time (participants were shown the front of each card and asked to guess if a given symbol would appear on the back before it was turned over). The experiment then picked four cards from the deck and participants were then presented with Wason's (1968) selection task (see §1.1). The results provided some evidence for the influence of experience (the learning phase) on people's judgements in the selection task.

Oaksford and Wakefield (2003) used a similar method of sequential exposure to materials which provided participants with contextual data over time before they completed a selection task. This approach initially provided frequency information about the number of cards with given labels in a large pack directly. They then presented participants with example cards, one at a time, which had labels in proportion with those provided in the initial frequency information. Participants were asked about each card they were shown and then asked a key experimental selection task question. This approach was intended to emulate the natural sampling through which people normally acquire information about relative frequencies. Gigerenzer and Hoffrage (1995) had shown that people were better at Bayesian probability problems where they were given information about frequency (e.g. "8 of every 10 women with breast cancer will get a positive mammography") rather than a probability (e.g. "If a woman has breast cancer, the probability is 80% that she will get a positive mammography"). Oaksford and Wakefield (2003) showed that sequential exposure to materials (which allowed the development of an understanding of relative frequencies) led to responses to selection task questions that better reflected the probability model based on the frequencies provided to participants than had been found by Oberauer et al. (1999) which had provided only narrative information. Using a similar approach with sequential presentation, Oaksford and Moussakowski (2004) found that people

responded to a selection task in a way that better reflected the frequencies provided than Yama (2001) had observed (Yama, 2001, had also only provided narrative information).

This experiment intends to manipulate the frequency with which participants will anticipate given events occurring. Based on these previous results showing that people's responses appear better informed by frequency data when it is provided sequentially, the experiment will include a learning phase. This phase will provide information about individual instances relevant to the conditional to be considered one at a time.

Participants will be asked questions about the information they are shown to support engagement with it.

#### **4.2.2. Predictions**

In this experiment, participants will be provided with information about the proportion of animals of each of three different types and colours. This information will be given both directly in the form of percentages and indirectly through a learning phase which will show 50 pictures of animals of the relevant types and colours. The information would suggest that an MP inference is much more likely to be true when a "not" negated second premise is used than when a contrast class member is used in its place. The proportions given were based on those suggested by Oaksford and Chater (2007), modified to allow the context to be presented in 50 pictures. The proportion of animals of each type and colour shown to participants is detailed in table 4.2.

*Table 4.2 – Contextual information provided about animals*

		Type of animal			Total
		Cat	Dog	Rabbit	
Colour of animal	Black	30% (15)	10% (5)	0% (0)	40% (20)
	White	6% (3)	2% (1)	22% (11)	30% (15)
	Brown	4% (2)	4% (2)	22% (11)	30% (15)
Total		40% (20)	16% (8)	44% (22)	100% (50)

Percentage of given type and colour shown to the participants (number of given type and colour presented in learning phase)

Participants would be asked to use this information to test the rule "If it is not a cat, then it is not black". They would be presented with the rule as a major premise and a suitable statement for the minor premise. For the MP inference, the minor premise would be



“not a cat” in the “not” negated condition and “a dog” in the contrast class condition. Participants would then be asked whether they would conclude either that “The animal is not black” (which is equivalent to endorsement of MP in this example) or that “The animal is black”.

The scenarios presented would suggest that animals of the relevant types and colours exist in such a probability of a MP inference being correct when “not” negation is used in the second premise is 83.33%. However, when the contrast class member ‘dog’ is used in the second premise, this probability drops to 37.5%. Therefore, given this context, it is expected that participants make fewer MP inferences when the contrast class member ‘dog’ is used in the second premise in place of ‘not a cat’. Similarly, confidence ratings that the inference is correct are anticipated to be lower for MP inference when the contrast class is used than the explicit negation.

Conversely, no significant differences are predicted with AC inference between the contrast class and explicitly negated condition for either confidence ratings or number of inferences made. This is because the scenarios presented only provide a small difference between the probabilities that AC inference is correct when “not” negation is used (83.33%) and when the contrast class is used (80%).

### **4.2.3. Method**

#### **4.2.3.1. Design**

This experiment used a within-participants design with inference type (MP, DA, AC and MT), conclusion polarity (“not” negated and contrast class for MP and AC inference conditions, affirmative for the other conditions) and order (before learning phase and after learning phase) as factors. The dependent variables were responses to the inference tasks (whether they made the inference or not) and their stated confidence in their response (provided using a slider on a visual analogue scale which was assigned a response between 1 and 100).

#### **4.2.3.2. Participants**

Participants were recruited through the ‘Amazon Mechanical Turk’ system to complete the experimental survey. All people who completed the survey (including a small number of responses which were excluded from the data because they may have been

duplicates – either sharing a Mechanical Turk ID or IP address) received a small payment (between US\$0.50 and US\$1.00). After these exclusions, the sample consisted of 272 people, was 52% female and was aged between 18 and 75 (median age 31). A total of 97% of participants reported that English was their first language.

#### 4.2.3.3. Materials

A scenario was prepared in which a vet was testing the following conditional rule in respect to the animals that she saw:

*If it is not a cat, then it is not black*

Six inference tasks were developed using this rule. These took the following form:

*The vet is considering the following rule about the animals that she sees:*

- *If it is not a cat, then it is not black.*

*The vet is told that the next animal she will see is:*

[One of the following minor premises was provided for each question]

- *not a cat.* [MP with “not” negation]
- *a dog.* [MP with contrast class]
- *a cat.* [DA]
- *not black.* [AC with “not” negation]
- *white.* [AC with contrast class]
- *black.* [MT]

*Please select the option below that best describes what she should conclude about the next animal.*

[One of the following pairs of options was provided with the options in random order for each question]

[For MP and DA questions:]

*That the animal is not black*

*That the animal is black*

[For AC and MT questions:]

*That the animal is not a cat*

*That the animal is a cat*

In addition, 50 questions were prepared to form the learning phase. These questions each consisted of a photograph of an animal of a given type and colour. Photographs of relevant animals against a white background were found which had no features apart from the animal. The photographs were cropped and reduced in size to ensure they would all display at the same size in an online survey and that the photo and questions on each learning phase page would fit on a single screen at typical resolution. These photos were each combined with multiple choice questions about the photo which participants had to answer to move on. This was intended to ensure that participants attended to the materials. The following questions were used:

*What type of animal is this?* [Answer options: *Dog; Cat; Rabbit*]

*What colour best describes this animal?* [Answer options: *Black; White; Brown*]

Full details of the scenario used in this experiment and questions asked are included in appendix 3 (see §A3.1).

#### 4.2.3.4. Procedure

Participants were directed to the web-based survey (using [www.surveymoz.com](http://www.surveymoz.com)) from the Amazon Mechanical Turk System ([www.mturk.com](http://www.mturk.com)). They saw an information screen and had to confirm consent to proceed. They then provided basic demographic information before being provided with instructions for the first part of the experiment.

The first part of the experiment provided participants details of the proportion of animals that the vet sees of different types (cats, dogs, rabbits) and colours (black, white, brown). These were provided as the percentages given in table 4.2 above. It then asked the six inference questions for the first time. These inference questions were presented one to a page in random order. Participants had to provide a response to the inference question and click on the slider bar to indicate their confidence in the response they had given before they were able to move onto the next page. The slider bar was

labelled with 'Not at all confident' at one end and 'Completely confident' at the other. Responses were recorded at a number between 1 and 100 based on how far down the slider bar the participant clicked.

The participants were then given instructions for the learning phase. The learning phase consisted of fifty pages. Each page had a picture of an animal and asked two multiple choice questions about the animals type and colour. Participants had to answer both questions on each page before they could continue. Learning phase pages were presented in a random order.

Participants were then given instructions for the second set of inference questions. This stage used the same six inference questions that participants had answered earlier in random order. However, the participants were not given any information about the proportion of animals on the instructions page.

Finally, participants were presented with a page that contained the verification task. This presented nine response boxes in a three by three grid labelled animal type (cat, dog, rabbit) on one axis and colour (black, white, brown) on the other. Participants were instructed to enter how many of the next 100 animals that the vet would see would be in each category. If participants attempted to proceed without their responses summing to 100, they were returned to this page with an instruction providing the total value they had entered and asking them to make sure their responses added up to 100.

A final page provided participants with a code to enter on the Mechanical Turk system to confirm that they had completed the survey, thanked them for their time and provided contact information if they had any questions.

#### **4.2.4. Results**

Raw responses to confidence questions were numbers between 1 and 100 reflecting the extent to which the participant considered their response to the inference question were correct. These were converted into scores reflecting the participants' confidence that inferences were correct based on whether they made the inference in responding to the inference question or not. Where a response to the inference question was consistent with the relevant inference, the relevant response to the confidence question was simply

divided by 100 to provide a proportion. Where a response to the inference question was inconsistent with the relevant inference, the relevant response to the confidence question was simply divided by 100 and subtracted from 1 to provide the proportion confidence that that inference was correct.

The proportion of responses in line with the relevant inferences and average confidence scores are set out in table 4.3.

*Table 4.3 – Summary of responses to chapter 4, experiment 1*

	Trials before learning phase					
	MP		DA	AC		MT
	Not	Contrast		Not	Contrast	
Proportion making Inference	85.66%	83.09%	83.82%	86.40%	81.62%	86.03%
Confidence of Inference (standard deviation)	.70 (0.28)	.68 (0.28)	.58 (0.29)	.66 (0.28)	.59 (0.29)	.66 (0.28)
	Trials following learning phase					
	MP		DA	AC		MT
	Not	Contrast		Not	Contrast	
Proportion making Inference	90.07%	83.46%	89.34%	93.01%	87.13%	90.81%
Confidence of Inference (standard deviation)	.69 (0.27)	.61 (0.30)	.61 (0.29)	.65 (0.28)	.62 (0.28)	.67 (0.26)
<i>Proportion making inference as a percentage to 2 decimal places</i>						
<i>Mean Confidence (standard deviation) to 2 decimal places</i>						

Planned comparisons using McNemar  $\chi^2$  tests were used to compare the number of inference question responses consistent with the relevant inference in the “not” negated and contrast class conditions for the MP and AC inferences, before and after the learning phase. Neither inference showed a significant difference between “not” negated and contrast class conditions before the learning phase. After the learning phase, significant differences between the number of inferences made in “not” negated and contrast class conditions were observed for both MP ( $\chi^2(1) = 6.000$ ,  $p=.021$ ) and AC ( $\chi^2(1) = 5.818$ ,  $p=.024$ ) inferences.

Difference between confidence scores was analysed using a 2x2x2 repeated measures ANOVA with order (before and after the learning phase), inference (MP and AC) and negation used (“not” negation and contrast class) as factors. This found significant main effects of inference ( $F(1,271) = 9.600, p=.002, \eta_p^2 = .034$ ) and negation ( $F(1,271) = 22.104, p<.001, \eta_p^2 = .075$ ), a significant two-way interaction between order and inference ( $F(1,271) = 7.205, p=.008, \eta_p^2 = .026$ ) and a significant three-way interaction between each of the factors ( $F(1,271) = 7.801, p=.006, \eta_p^2 = .028$ ).

Four planned t-test comparisons (with Bonferroni adjusted significance levels of  $p=.013$ ) were used to compare confidence scores in the “not” negated and contrast class conditions. There were done for the MP inference before the learning phase which was not significant and the AC inference before the learning phase which was significant ( $t(271) = 3.048, p=.003$ ). After the learning phase, t-tests were used to compare the responses in the MP inference which showed significant differences between the “not” negated and contrast class conditions ( $t(271) = 4.608, p<.001$ ) and AC inference which did not reach the required Bonferroni level of significance ( $t(271) = 1.854, p=.065$ ).

#### 4.2.4.1. Engagement with Learning Phase

The experimental survey provided two ways of monitoring engagement with the learning phases. The first was the responses to the questions that people answered to describe the pictures they saw in the learning phase which was intended to ensure they paid attention. The second was the final question which indicated the extent to which participants had integrated an understanding of the relative frequencies presented.

Participants were asked two questions about each of the 50 pictures they see (about the type and colour of the animal): a total of 100 learning phase questions. Errors per participant ranged from 0 to 33 (with a median of 0, mean of 0.7 and standard deviation of 2.26). Only 96 participants (35%) made any errors and only three participants (about 1%) made more than three errors. It would, therefore, appear that people responded with reasonable accuracy throughout the learning phase. However, that does not say anything about whether they attended sufficiently to learn anything about the relative frequencies presented.

The final question in the experimental survey asked participants to provide their own estimates of the number of animals of each type and colour. The coefficient of determination (the  $R^2$  statistic) was calculated for each participant based on how well their responses to this question matched the model implied by the information provided in the first set of instructions and by the learning stage. The  $R^2$  statistic represents the proportion of variation in a variable (in this case participants responses to the validation task) that is accounted for by another variable (in this case the responses predicted by the frequencies presented in the learning task). For each participant, it is calculated by dividing the residual sum of squares (the sum of the squared differences between each response and its prediction) by the total sum of squares (the sum of squared differences between each response and the participant's mean response) and subtracting this number from 1. A result of between 0 and 1 represents the extent to which the model data (in this case the responses predicted by learning task frequencies) predicts the response data. A score of 0 represents response data that is not at all predicated by the response data. A score of less than 0 (which occurs where the residual sum of squares is higher than the total sum of squares) occurs where the predicted responses are less useful a prediction than predicting the same value for each response. This statistic is often used to confirm the extent to which a model fits response data (e.g. Oaksford & Chater, 2007).

The  $R^2$  statistics for each participant ranged between -1,070 and 1, had a median of 0.019, a mean of -17.85 and a standard deviation of 126.50. This suggests that, for a large number of participants, the information presented in the first set of instructions and the learning phase provides a very poor model for their understanding of the number of animals of each type and colour.

To consider only participants who demonstrated some impact of the learning phase in response to the final question, all those with an  $R^2$  score of greater than zero were selected to form a sub-sample. This smaller sample consisted of 139 participants.

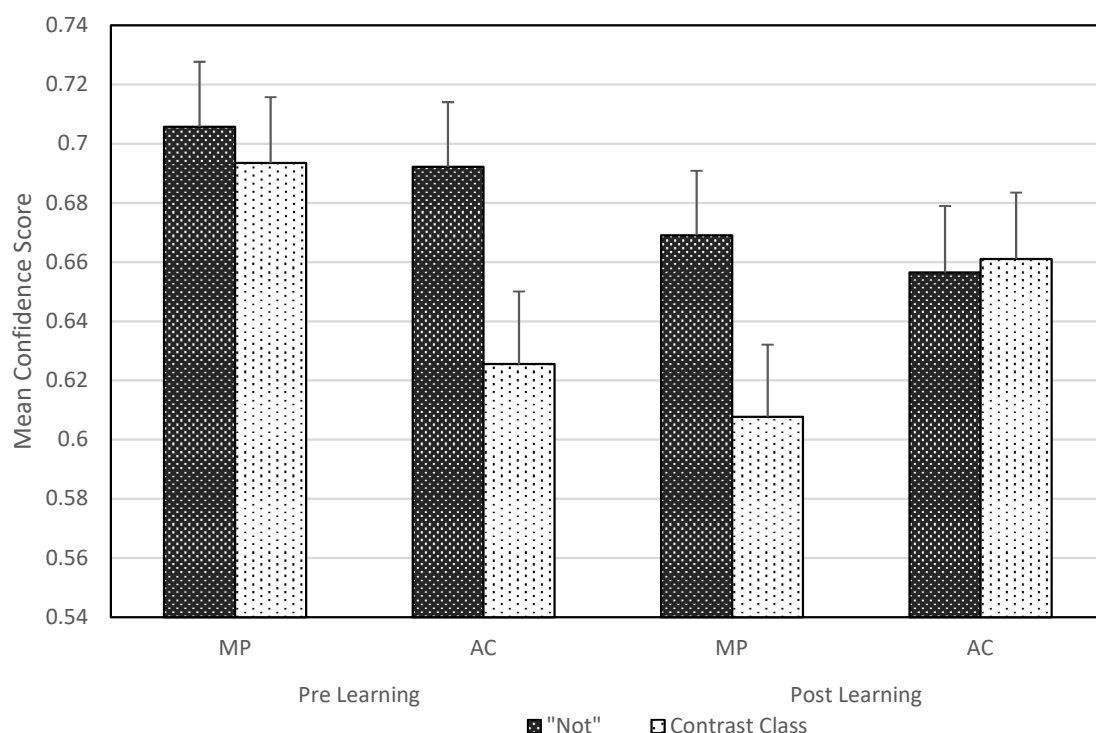
The previous planned comparisons were repeated using this narrower sample whose estimates of the frequency of animal types and colours showed evidence of being informed by the learning phase. McNemar  $\chi^2$  tests were again used to compare the number of inference question responses consistent with the relevant inference in the

“not” negated and contrast class conditions for the MP and AC inferences, before and after the learning phase. Neither inference showed a significant difference between “not” negated and contrast class conditions before the learning phase. After the learning phase, a significant difference between the number of inferences made in “not” negated and contrast class conditions was observed for MP ( $\chi^2(1) = 11.636, p=.001$ ) inferences but not AC inferences. In this narrower sample, participants endorsed 95.0% of MP inferences when “not” was used and only 83.5% when the contrast class was used. AC inferences were endorsed 95.7% of the time when “not” negation was used and 92.8% of the time in the contrast class condition.

A 2x2x2 repeated measures ANOVA was also conducted on this smaller sample’s confidence scores with order (before and after the learning phase), inference (MP and AC) and negation used (“not” negation and contrast class) as factors. This found significant main effects of order ( $F(1,138) = 4.481, p=.036, \eta_p^2 = .031$ ) and negation ( $F(1,138) = 6.315, p=.013, \eta_p^2 = .044$ ), a significant two-way interaction between order and inference ( $F(1,138) = 7.472, p=.007, \eta_p^2 = .051$ ) and a significant three-way interaction between each of the factors ( $F(1,138) = 10.799, p=.001, \eta_p^2 = .073$ ).

Planned comparisons using t-tests (with Bonferroni adjusted significance levels of  $p=.013$ ) were used to compare confidence scores in the “not” negated and contrast class conditions for the MP and AC inferences. Before the learning phase, a significant difference was found between confidence scores for AC inferences ( $t(138) = 2.586, p=.011$ ) but not MP inferences. After the learning phase, significant differences were found for MP ( $t(138) = 3.104, p=.002$ ) but not AC inferences. The mean confidence levels for these inferences are shown in figure 4.1 (error bars represent standard error).





*Figure 4.1* – Chart Showing Mean Confidence Scores for MP and AC Inferences of Participants that Demonstrated Learning in Experiment 1.

#### 4.2.4.2. Participants' Probability Estimates

The final question, in which participants provided their estimates of the number of animals of each type and colour, also allows us to estimate internal probabilities that each inference is correct given the premises provided. We can, therefore, look at whether people's confidence that an inference is true is correlated with the probability that it is true based on their internal understanding of the frequency of different occurrences.

Pearson correlation tests were used to compare confidence scores for each of the inference questions to the conditional probability that the inference is correct based on the participant's self-report of animal frequencies. For the inference tests before the learning phase, a significant correlation was only found between the confidence in the inference and the probability based on self-reported animal frequencies for the MP inference using the contrast class ( $r=.171$ ,  $p=.005$ ). For the inference tests following the learning phase, significant correlations were found between participants confidence scores and the conditional probabilities based on their self-reported understanding of

animal frequencies across most conditions. This includes MP inferences using “not” negation ( $r=.153$ ,  $p=.011$ ) and the contrast class ( $r=.278$ ,  $p<.001$ ), DA inferences using “not” ( $r=.134$ ,  $p=.027$ ) and AC inferences using the contrast class ( $r=.179$ ,  $p=.003$ ). Post-learning, significant correlations were only not found for AC inferences using “not” and MT inferences.

#### **4.2.5. Discussion**

This experiment has shown that, where participants can demonstrate some learning of the frequency data, the use of implicit negation suppresses MP inference in the manner predicted. In contrast, AC inferences are not suppressed by the use of implicit negation based on this frequency data as would be expected if participants inferences are motivated based on the relevant conditional probabilities.

This provides some initial support for Oaksford and Chater’s (2007) intuition about how the conditional probability model could be used to motivate inferences where implicit negations from the contrast class are used. While these results are consistent with their theory, other approaches cannot be ruled out based on this experiment. MP inferences were suppressed when the low probability contrast class member was used as an implicit negator, but we cannot be sure whether this was because of the frequency data provided or because of a more general implicit negation effect when considering MP inferences. For example, a mental models account (e.g. Johnson-Laird & Byrne, 2002) could argue that this effect was due to the use of implicit negation in MP making consideration of the task more complex (requiring an additional step to form the mental models required). The next experiment will address this by using MP inferences both where the frequency data would suggest that “not” negated and contrast class negated conclusions should be endorsed at a similar level and where frequency data is expected to motivate a difference.

When data from all those that completed the experiment was taken into account (including those that showed no evidence of learning about the frequencies) a significant suppression effect was also seen when the contrast class member was used to provide implicit negation in the AC task as well as the MP task. This is not what was predicted and may suggest that there is an implicit negation effect independent of the

conditional probabilities (but a smaller effect that disappears when the frequency data has been understood).

While MP inferences were suppressed in the contrast class condition, they were still relatively high. It should be noted that while the relevant conditional probability was .37, it was never predicted that inference levels would fall this far. The conditional probability model proposed that the relevant conditional probabilities inform the likely level of inferences, but the levels are not likely to be identical. This is particularly the case with MP which people are normally ready to endorse.

Further support for the conditional probability model comes from the strong correlations between participants internalised understanding of frequencies (as reported in the final task) and their confidence scores in relation to most inferences. This is what would be expected if people had developed their own understanding of the conditional probabilities based on the learning phase and this information informed their inference responses.

### **4.3. Experiment 2**

The previous experiment showed that MP would show an implicit negation effect – suppressing inference levels – as Oaksford and Chater (2007) predicted when a contrast class member was used in the premises. However, this result is not sufficient to demonstrate that it is the conditional probabilities that have led to this effect and rule out other theories.

This experiment will, therefore, seek to expand the result. Firstly, it will look at whether frequencies can be modified to both create an implicit negation effect for MP and remove that effect in different conditions. Secondly, it will see if a similar effect can be induced in a different type of inference, AC. The third key change will be to use more extreme differences in frequency levels to see if this increases the suppression effect compared to that seen in the prior experiment. If an effect for MP and AC can be created and removed in different frequency conditions, with the effect increased through the use of more extreme frequency differences, this will provide substantial support for the conditional probability model. As discussed above (§4.1.3) it is not clear how mental models or heuristic-analytic approaches could explain such effects.

This experiment will use a second scenario alongside the one used in the previous experiment to test whether the effects can be replicated with different scenarios. The learning test will be shortened to allow participants to complete the tasks using both scenarios in a reasonable amount of time.

One potential criticism of the presentation of the previous experiment is that the presentation of direct frequency data at the beginning may have primed participants to take a probabilistic approach. It is not clear based on responses to the inference questions before the learning phase that they were taking this in, but it will be removed in this experiment to prevent any risk of priming. Because this has been removed – and also to avoid priming people by considering the inference questions before the learning phase – there is no initial set of inference tests.

#### **4.3.1. Predictions**

In this experiment, participants will complete two learning tasks, each showing 30 pictures and each followed by a set of inference questions. One learning task and set of inference questions will use the same scenario presented in experiment 1. The other learning task and set of inference task will use a new scenario based on vehicles of different types and colours. In this latter scenario, participants will be asked to test the rule “If it is not white, then it is not a van”. They would be presented with the rule as a major premise and a suitable statement for the minor premise. For the AC inference, the minor premise would be “not a van” in the “not” negated condition and “a motorbike” in the contrast class condition. Participants would then be asked whether they would conclude whether “the vehicle is not white” (equivalent to endorsing the AC inference) or “The vehicle is white”.

The proportions of vehicles or animals of different types and colours shown in each learning phases have been manipulated to suggest that, for one inference type, the inference is much more likely be true when a “not” negated second premise is used than when a contrast class member is used in its place. This is, again, based on the approach proposed by Oaksford and Chater (2007). Each participant will see one scenario intended to suggest that MP inferences are more likely when “not” negation is used in place of the contrast class and no difference in AC inferences between these conditions.

The other scenario will keep the likelihood of MP inferences constant between “not” negated and contrast class conditions while suggesting that AC is more likely where “not” negation is used than the contrast class. Participants will be randomly assigned to an animal scenario that varies the probabilities associated with either MP or AC and the vehicle scenario will then vary the probabilities associated with the other.

*Table 4.4 – Contextual information provided – MP manipulation*

		Colour of vehicle / type of animal			Total
		White / Cat	Blue / Dog	Red / Rabbit	
Type of	Van / Black	8	2	0	10
vehicle /	Motorbike /	0	0	10	10
colour of	White				
animal	Car / Brown	0	0	10	10
Total		8	2	20	30

Number of given type and colour presented in learning phase

For the scenario where MP is varied the number of vehicles or animals of different types and colours presented in the learning phase (see table 4.4) suggests that the probability of a MP inference being correct when “not” negation is used in the minor premise is 91%. However, when a contrast class member (‘blue’ or ‘dog’) is used in the minor premise, this probability drops to 0%. Therefore, given this context, it is expected that participants make fewer MP inferences when the contrast class member is used in the minor premise instead of the “not” negated premise. For the same reason, confidence ratings are expected to be lower when drawing the MP inference using the contrast class member in the minor premise than when explicit negation is used.

Where AC inferences are tested in this scenario, no significant differences are anticipated in the level of inferences made or confidence ratings in those inferences between the contrast class and explicitly negated condition. The scenarios presented provide no difference between the probabilities that AC inference is correct when “not” negation is used (100%) and when the contrast class is used (100%).

*Table 4.5– Contextual information provided – AC manipulation*

		Colour of vehicle / type of animal			Total
		White / Cat	Blue / Dog	Red / Rabbit	
Type of	Van / Black	8	0	0	8
vehicle /	Motorbike /	2	0	0	2
colour of	White				
animal	Car / Brown	0	10	10	20
Total		10	10	10	30

Number of given type and colour presented in learning phase

For the scenario where AC is varied, the number of vehicles or animals of different types and colours presented in the learning phase (see table 4.5) suggests that the probability of an AC inference being correct when “not” negation is used in the minor premise is 91%. However, when a contrast class member (‘white’ in reference to animals or ‘motorbike’ for vehicles) is used in the minor premise, this probability drops to 0%. Therefore, given this context, it is expected that participants make fewer AC inferences when the contrast class member is used in the minor premise instead of the “not” negated premise. For the same reason, confidence ratings are expected to be lower when drawing the AC inference using the contrast class member in the minor premise than when explicit negation is used.

Where MP inferences are tested in this scenario, no significant differences are anticipated in the level of inferences made or confidence ratings in those inferences between the contrast class and explicitly negated condition. The scenarios presented provide no difference between the probabilities that MP inference is correct when “not” negation is used (100%) and when the contrast class is used (100%).

### **4.3.2. Method**

#### **4.3.2.1. Design**

A mixed design was used for this experiment. Inference type (MP, DA, AC and MT), conclusion polarity (“not” negated and contrast class for MP and AC inference types and only affirmative for DA and MT types) and manipulation (MP manipulation or AC manipulation) were within participants factors. Scenario was a between participants factor with each participant being randomly assigned to one scenario for the MP manipulation condition and the other for the AC manipulation condition. All

participants, therefore, responded to all inference types and conclusion polarities in both manipulation conditions. They also saw both scenarios. However, roughly half the participants saw the animal scenario in the MP manipulation condition and the vehicle scenario in the AC manipulation condition. For the other participants, the scenarios were reversed.

Responses to the inference tasks (whether they made the inference or not) and confidence in that response (based on a visual analogue scale using a slider providing a response between 1 and 100) were dependent variables.

#### 4.3.2.2. Participants

A total of 334 unique participants completed the experimental survey after being recruited through the ‘Amazon Mechanical Turk’ system. This total is the number left after a small number of exclusions to ensure each participant was unique (responses that either shared a Mechanical Turk ID or IP address). All people completing the survey received a small payment (US\$0.50). Participants were aged between 18 and 83 (median age 36), 53.6% were female and 96.4% reported English as their first language.

#### 4.3.2.3. Materials

The scenario developed for the previous experiment involving animals was reused. Two sets of 30 learning phase tasks were used for this experiment with images of animals of different types and colours to match the frequencies for the MP variation and AC variation conditions as set out above. The form of the learning and inference tasks were otherwise the same as the previous experiment.

In addition, a new scenario was developed in which a police traffic officer was testing the following conditional rule concerning the vehicles that she saw:

*If it is not white, then it is not a van*

As with the previous experiment, six inference tasks were generated which used this rule as follows:

*The police traffic officer is considering the following rule about the vehicles that she sees:*

- *If it is not white, then it is not a van*

*A colleague up the road radios the officer to tell her that the next vehicle is:*

[One of the following minor premises was provided for each question]

- *not white.* [MP with “not” negation]
- *blue.* [MP with contrast class]
- *white.* [DA]
- *not a van.* [AC with “not” negation]
- *a motorbike.* [AC with contrast class]
- *a van.* [MT]

*Please select the option below that best describes what she should conclude about the next vehicle.*

[One of the following pairs of options was provided with the options in random order for each question]

[For MP and DA questions:]

*That the vehicle is not a van*

*That the vehicle is a van*

[For AC and MT questions:]

*That the vehicle is not white*

*That the vehicle is white*

As with the animal scenario, two sets of 30 learning phase tasks were generated using this scenario for this experiment. These had images of vehicles of different types and colours to match the frequencies for the MP variation and AC variation conditions as set out above. Appropriate photographs showing vehicles of the required types and colours against a white background with no other features were identified. The photographs were cropped and resized to appear the same size in the online survey and ensure the picture and questions would appear on a single page at typical resolutions. These questions used beneath each picture for each learning phase task were:



*What type of vehicle is this?* [Answer options: *Van*; *Motorbike*; *Car*]

*What colour is this vehicle?* [Answer options: *White*; *Blue*; *Red*]

Full details of both scenarios used in this experiment and questions asked are included in appendix 3 (see §A3.1 for the ‘animals’ scenario and §A3.2 for the ‘vehicles’ scenario).

#### 4.3.2.4. Procedure

Participants were recruited through the Amazon Mechanical Turk System and completed the experiment using a web-based survey (using [www.surveymzmo.com](http://www.surveymzmo.com)). An information screen requested that they confirm consent before they were able to proceed. As this stage, the survey randomly assigned each participant to randomly see the animal scenario in either the form that varied the likelihood of MP inferences or the form that varied the likelihood of AC inferences. They were then allocated to the form of vehicle scenario that varied the likelihood of the other inference type. The survey also randomly allocated the participant to see one of the two scenarios first and the other second. The following screen sought basic demographic information before instructions for the first learning phase were provided.

Each learning phase consisted of thirty pages presented in random order. Each page had a picture of a vehicle or animal and asked participants what type of vehicle or animal was shown (with multiple choice answers *car*, *van* and *motorbike* presented in random order for the vehicle scenario and *cat*, *dog* and *rabbit* presented in random order for the animal scenario) and what colour the vehicle or animal was (with possible responses *white*, *blue* and *red* for vehicles and *black*, *white* and *brown* for animals also presented in random order). Participants had to answer both questions before continuing to the next page.

After each learning phase, participants were then given instructions for the inference questions. The six inference questions for each scenario (as described in §4.2.3.3 and §4.3.2.3) were then presented in random order with one inference on each page. Participants had to provide a response to the inference question and click on a slider bar to indicate their confidence in the response they have given before they were able to move onto the next page. The slider bar was labelled with ‘Not at all confident’ at one

end and ‘Completely confident’ at the other. Responses were recorded as a number between 1 and 100 based on how far down the slider bar the participant clicked.

After each learning phase and set of inference questions, participants were asked to complete a verification task. This task was presented on a single page with a three by three grid of response boxes. The columns were labelled with the possible colours used in the scenario (*white, blue, red* for vehicles and *black, white, brown* for animals) and the rows with types (*car, van, motorbike* for vehicles and *cat, dog, rabbit* for animals). The order of the columns and rows was randomised for each participant. Participants were asked to enter how many of the next 100 vehicles that the police traffic officer sees would be in each category or how many of the next 100 animals that the vet sees would be in each category. If participants attempted to proceed without their responses summing to 100, they were returned to this page with an instruction providing the total value they had entered and asking them to make sure their responses added up to 100.

After completing one set of learning, inference and verification tasks, participants started the tasks for the other scenario.

A final page provided participants with a code to enter on the Mechanical Turk system to confirm that they had completed the survey, thanked them for their time and provided contact information if they had any questions.

#### **4.3.3. Results**

As with the prior experiment, responses to confidence questions were converted into scores reflecting the confidence the inference tested was correct. Table 4.6 sets out the proportion of responses in line with the relevant inferences and average confidence scores.

Table 4.6 – Summary of responses to chapter 4, experiment 2

	Trials in MP manipulation condition					
	MP		DA	AC		MT
	Not	Contrast		Not	Contrast	
Proportion making Inference	97.3%	68.6%	93.1%	94.0%	95.8%	90.7%
Confidence of Inference (standard deviation)	.75 (0.26)	.53 (0.35)	.78 (0.26)	.76 (0.27)	.79 (0.25)	.71 (0.28)
	Trials in AC manipulation condition					
	MP		DA	AC		MT
	Not	Contrast		Not	Contrast	
Proportion making Inference	93.7%	94.3%	91.9%	93.4%	55.1%	92.8%
Confidence of Inference (standard deviation)	.81 (0.25)	.83 (0.22)	.68 (0.28)	.70 (0.29)	.48 (0.34)	.81 (0.26)
Proportion making inference as a percentage to 2 decimal places						
Mean Confidence (standard deviation) to 2 decimal places						

For the MP manipulation trials, McNemar  $\chi^2$  tests were used for a planned comparison of the number of inference question responses consistent with the relevant inference in the “not” negated and contrast class conditions for the MP and AC inferences. These found a significant difference between the number of inferences made in “not” negated and contrast class conditions for MP inferences ( $\chi^2(1) = 85.333$ ,  $p < .001$ ) but not AC inferences. The same tests were done for the AC manipulation trials and found significant differences between the responses to the “not” negated and contrast class conditions for AC inferences ( $\chi^2(1) = 110.703$ ,  $p < .001$ ) but not MP inferences.

Difference between confidence scores was analysed using a 2x2x2x2 mixed measures ANOVA. The repeated measures factors were manipulation (MP manipulated and AC manipulated), inference (MP and AC) and negation used (“not” negation and contrast class). The independent measure factor was scenario (animal scenario in MP manipulation and animal scenario in AC manipulation). This found a significant three

way interaction between manipulation, inference and negation type ( $F(1,332) = 147.898, p < .001, \eta_p^2 = .308$ ). It also found a significant two way interactions between manipulation and inference ( $F(1,332) = 93.429, p < .001, \eta_p^2 = .402$ ), between manipulation and scenario ( $F(1, 332) = 12.934, p < .001, \eta_p^2 = .037$ ) and between negation and scenario ( $F(1,332) = 6.058, p = .014, \eta_p^2 = .018$ ). Significant main effects were found for inference ( $F(1,332) = 23.180, p < .001, \eta_p^2 = .065$ ), negation type ( $F(1, 332) = 93.429, p < .001, \eta_p^2 = .220$ ) and scenarios ( $F(1,332) = 9.482, p = .002, \eta_p^2 = .028$ ).

Four planned t-test comparisons (with Bonferroni adjusted significance levels of  $p = .013$ ) were used to compare confidence scores between the “not” negated and contrast class conditions for the MP and AC inferences in each of the MP manipulation and the AC manipulation conditions. In the MP manipulation condition, the MP inference showed a significant difference between the negation type conditions ( $t(333) = 10.325, p < .001$ ) but the AC inference failed to reach the necessary level of significance ( $t(333) = -1.971, p = .050$ ). In the AC manipulation condition, the difference between the negation type conditions failed to reach the required significance level for MP inference ( $t(333) = -2.050, p = .041$ ) but a significant difference was observed in the AC inferences ( $t(333) = 10.115, p < .001$ ).

#### 4.3.3.1. Engagement with Learning Phase

As with the previous experiment, only about 35% of participants made any mistakes at all on any learning task questions (118 participants) across either learning phase. Each participant answered 120 questions across both learning phases. The total number of mistakes made by each participant across both phases ranged from 0 to 74, with a median of 0, a mean of 0.8 and a standard deviation of 4.14.

As before, the final question relating to each scenario asked participants to provide their estimates of the number of animals or vehicles of each type and colour. Two  $R^2$  statistics could be calculated for each participant – one for each scenario – reflecting how well their responses to this question matched the model implied by the information provided in the learning stage (the  $R^2$  statistic is described in §4.2.4.1). These  $R^2$  statistics ranged from -2,125 to 0.999 with a mean of -48.65 and standard deviation of 310.32 (median  $R^2$  was 0.669). As with the prior experiment, this suggests that many participants may not have internalised much of the information provided in the learning

phase. Removing participants' responses to a given scenario where their  $R^2$  score was less than zero provided a sub-sample of responses where there was some evidence that the participant based their estimates of the likelihood of different options within the scenario on the information in the learning phase. This smaller sample consisted of 237 sets of responses to scenarios in the MP manipulation condition and 247 sets of responses to scenarios in the AC manipulation condition. Before this exclusion, each participant (334) had provided a set of responses for each of the two manipulation conditions.

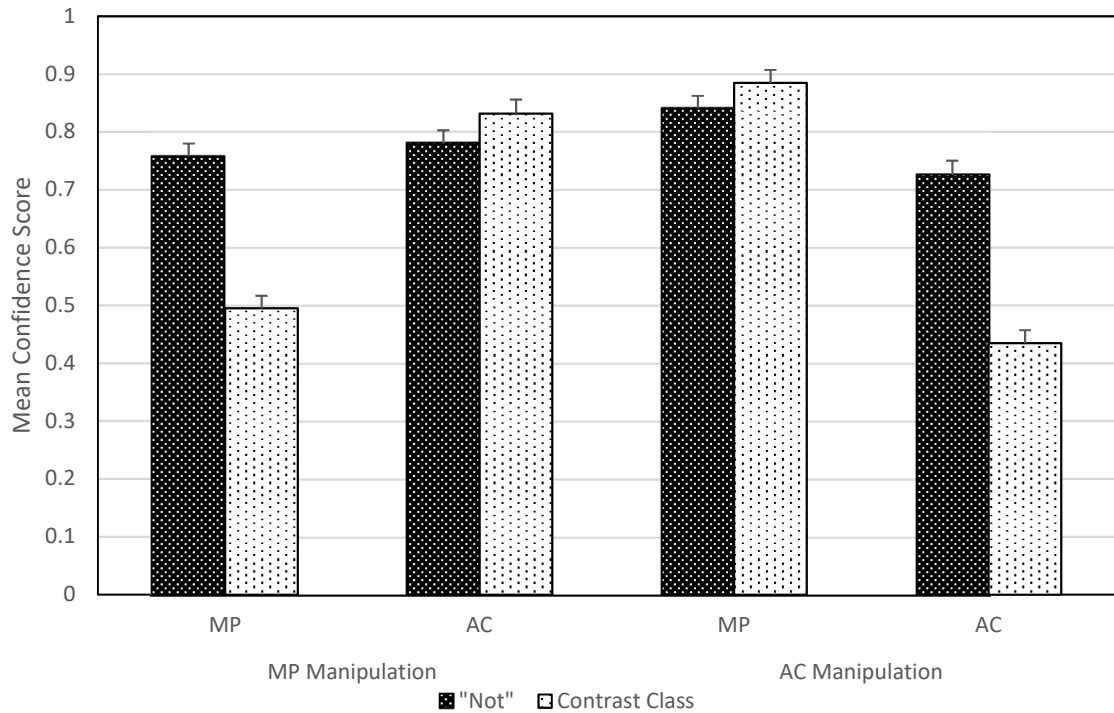
The same overall analysis was completed again with this smaller sample. Because not every participant had provided a response which was still included in both manipulation conditions, manipulation was treated as a between participants variable.

McNemar  $\chi^2$  tests were again used to compare the number of inference question responses consistent with the relevant inference in the "not" negated and contrast class conditions for the MP and AC inferences within each of the MP manipulation and AC manipulation conditions. Within the MP manipulation condition, a significant difference was again found between the number of inferences made in "not" negated and contrast class conditions for MP inferences ( $\chi^2(1) = 72.737, p < .001$ ) but not AC inferences. With MP manipulated, MP inference levels were 98.3% and 64.6% and AC inference levels were 94.9% and 98.3% for the "not" and contrast class conditions respectively. Within the AC manipulation condition, a significant difference was found between the responses to the "not" negated and contrast class conditions for AC inferences ( $\chi^2(1) = 108.138, p < .001$ ) but not MP inferences. With AC manipulated, MP inference levels were 97.6% and 98.0% and AC inference levels were 96.8% and 51.4% for the "not" and contrast class conditions respectively. This mirrors the results found for the overall sample above.

Difference between confidence scores was analysed using 2x2x2x2 mixed measures ANOVA using this smaller sample. For this test, the repeated measures factors were inference (MP and AC) and negation used ("not" negation and contrast class). The independent measures factors were manipulation (MP manipulated and AC manipulated) and scenario (animal scenario in MP manipulation and animal scenario in AC manipulation). This found significant three-way interactions between manipulation,

inference and negation type ( $F(1, 480) = 236.801, p < .001, \eta_p^2 = .330$ ) and between manipulation, inference and group ( $F(1, 480) = 5.588, p = .018, \eta_p^2 = .012$ ). It also found a significant two way interactions between manipulation and inference ( $F(1, 480) = 336.346, p < .001, \eta_p^2 = .412$ ) and between inference and group ( $F(1, 480) = 3.876, p = .050, \eta_p^2 = .008$ ). Significant main effects were found for inference ( $F(1, 480) = 14.225, p < .001, \eta_p^2 = .029$ ), negation type ( $F(1, 480) = 121.214, p < .001, \eta_p^2 = .202$ ) and group ( $F(1, 408) = 15.628, p < .001, \eta_p^2 = .032$ ). This varied from the results when a similar ANOVA was conducted using the overall sample in that the interactions between manipulation and group and between inference and negation were not significant. The three-way interaction between manipulation, inference and group and the two-way interaction between group and inference were also not observed in the full sample.

Again, four planned t-test comparisons (with Bonferroni adjusted significance levels of  $p = .013$ ) were used to compare confidence scores between the “not” negated and contrast class conditions for the MP and AC inferences in each of the MP manipulation and the AC manipulation conditions. In the MP manipulation condition, significant differences between the negation conditions were observed for both the MP inference ( $t(236) = 10.201, p < .001$ ) and the AC inference ( $t(236) = -3.133, p = .002$ ). In the AC manipulation condition, significant differences were also observed for both the MP inference ( $t(246) = -3.184, p = .002$ ) and the AC inferences ( $t(246) = 11.727, p < .001$ ). The mean confidence levels for these inferences is shown in figure 4.2 (error bars represent standard error).



*Figure 4.2 – Chart Showing Mean Confidence Scores for MP and AC Inferences of Participants that Demonstrated Learning in Experiment 2.*

#### 4.3.3.2. Materials Effects

Although a main effect of scenario was found and interactions between scenario and manipulation and between scenario and inference type, the size of these effects was small. Similarly, in the subsample including only responses to inference questions where the  $R^2$  based on the information in the relevant learning phase and participants' anticipated occurrence rates were greater than zero, the effects involving the group factor had low partial-eta-squared scores.

When the whole sample is considered, mean confidence scores were .70 for responses involving the animal scenario and .74 for responses involving the vehicle scenario. This reflects a pattern of results in which responses to the questions involving vehicles led to slightly higher confidence scores than the equivalent questions involving each inference, negation and manipulation condition except in one case. Confidence scores for questions involving AC inferences with contrast class negation in the AC manipulation condition had a slightly higher mean when animals were used (.84) than when vehicles were used (.82).

For each scenario in isolation, McNemar  $\chi^2$  tests were used to compare the number of inference question responses consistent with the relevant inference in the “not” negated and contrast class conditions for the MP and AC inferences within each of the MP manipulation and AC manipulation conditions. Within the MP manipulation condition, a significant difference was found between the number of inferences made in “not” negated and contrast class conditions for MP inferences for both the animal scenario ( $\chi^2(1) = 38.754, p < .001$ ) and the vehicle scenario ( $\chi^2(1) = 47.078, p < .001$ ) but not AC inferences for either scenario. Within the AC manipulation condition significant differences were found between the responses to the “not” negated and contrast class conditions for AC inferences for both the animal scenario ( $\chi^2(1) = 59.282, p < .001$ ) and the vehicle scenario ( $\chi^2(1) = 51.429, p < .001$ ) but not MP inferences for either scenario. This mirrors the results found for the overall sample above.

Confidence scores were compared for each scenario using two 2x2x2 mixed measures ANOVAs (one for the animal scenario and the other for the vehicles). These comparisons had inference (MP and AC) and negation used (“not” negation and contrast class) as repeated measures factors and manipulation (MP manipulated or AC manipulated) as an independent measures factor. Both found significant three-way interactions between manipulation, inference and negation type ( $F(1, 332) = 84.002, p < .001, \eta_p^2 = .225$  for animals and  $F(1, 332) = 96.520, p < .001, \eta_p^2 = .202$  for vehicles) and a significant two way interactions between inference and manipulation ( $F(1, 332) = 127.896, p < .001, \eta_p^2 = .278$  for animals and  $F(1, 332) = 156.878, p < .001, \eta_p^2 = .321$  for vehicles). They also both found significant main effects for negation ( $F(1, 332) = 48.047, p < .001, \eta_p^2 = .126$  for animals and  $F(1, 332) = 71.840, p < .001, \eta_p^2 = .178$  for vehicles) and manipulation ( $F(1, 332) = 6.796, p = .010, \eta_p^2 = .020$  for animals and  $F(1, 332) = 8.099, p < .001, \eta_p^2 = .024$  for vehicles). A significant main effect of inference type was only found when the vehicle scenario was used ( $F(1, 332) = 26.322, p < .001, \eta_p^2 = .073$ ). A marginally significant interaction between negation and manipulation was only found where the animal scenario was used ( $F(1, 332) = 4.331, p = .038, \eta_p^2 = .013$ ). As set out above, the overall results with both scenarios combined saw similar effects (including the main effect of inference not seen in responses to the animal scenario on their own). However, when both scenarios were looked at together, above, no main effect of manipulation was seen and nor was the interaction between negation and manipulation observed in the animal scenario alone identified.



As with the overall results, four planned t-test comparisons (with Bonferroni adjusted significance levels of .013) were used for each of the two scenarios. These compared confidence scores between the “not” negated and contrast class conditions for the MP and AC inferences in each of the MP manipulation and the AC manipulation conditions. In the MP manipulation condition, significant differences between the negation conditions were observed for the MP inferences in both using both the animal ( $t(163) = 6.461, p < .001$ ) and vehicle ( $t(169) = 8.103, p < .001$ ) scenarios but differences were not seen for AC inferences where either scenario was used. In the AC manipulation condition, significant differences were found for the AC inferences with both animal ( $t(169) = 7.770, p < .001$ ) and vehicle ( $t(163) = 6.500, p < .001$ ) scenario but not when MP inferences were used with either scenario.

#### **4.3.4. Discussion**

This experiment has shown that different frequency information can create and remove an implicit negation effect in MP and AC inferences based on the conditional probabilities implied. This effect appears larger than in the previous experiment, suggesting that more extreme differences in conditional probabilities can lead to greater differences in inference levels. This provides strong support for the conditional probability model of inference. It is also a result that is hard for other approaches to explain.

When the reduced sample that demonstrated learning from the learning phase was considered, the results were generally consistent with the overall result. One difference was that small significant differences in confidence scores were observed where they were not expected and where they had not been observed in the overall sample. There were small significant differences in confidence scores between MP inferences using “not” and the contrast class in the AC manipulation condition and also differences in confidence scores between AC inferences using “not” and the contrast class in the MP manipulation condition. These results could undermine the overall conclusion. However, these unexpected effects were much smaller than the predicted differences and in the opposite direction (see figure 4.2). They also only occurred when the confidence information was considered, they did not show up in tests related to the overall level of inferences. These confidence effects do not, therefore, appear to affect

the inferences that people are willing to draw. It may be that this is a separate small negation effect that is observed when materials of this type are used (people being slightly less confident in inferences that use “not” negated conclusions than ones using the contrast class). Any such small effect would appear to be minor compared to the effects when frequency information is manipulated which drives the main result of this experiment.

The fact that the overall results have better matched the predictions compared to those seen in the first experiment may reflect greater apparent engagement with the learning phase (70.5% of cases had  $R^2 > 0$  in this experiment compared to 51.1% in the previous experiment). It may be that the lack of direct frequency information up front and initial inference tasks led greater attention to that phase.

#### **4.4. Experiment 3**

This experiment will seek to show that the findings of the prior two experiments can be extended to DA and MT inferences. Demonstrating results consistent with the conditional probability model on a wider range of inferences will provide more support for the relevance of that model.

It is also possible that some accounts of reasoning are more appropriate to abstract tasks and materials while others are more appropriate to naturalistic tasks. The previous two experiments in this sequence have used naturalistic materials to generate and provide reasoning problems which leaves open the question of whether the anticipated effects based on the conditional probability model would also be seen when abstract materials are used. The current experiment will, therefore, test whether the use of naturalistic materials may be partly or wholly responsible for the effects observed on reasoning task responses in the prior experiments when the likelihood of contrast class members is manipulated. It will do this by using a set of abstract materials (asking people to reason over shapes of different colours) alongside the more natural vehicle scenario from experiment 2. If the approach to reasoning tasks differed when materials were abstract (and contrast class construction may be harder), then we would not expect to see the predicted effects where these materials are used.

#### **4.4.1. Predictions**

Like the previous experiment, participants will complete a series of tasks using two scenarios. As before, each scenario will consist of 30 learning tasks, six inference tasks and a verification task. One scenario will use the vehicle learning stage scenario from the previous experiment. The other scenario will use a new set of abstract learning stage scenario with shapes of different colours on them. The inference questions will use the rules, ‘If it is white, then it is a van’ following the vehicle learning phase and ‘If it is red, then it is a circle’ following the abstract scenario learning phase. DA and MT inferences would have both “not” negation (using “not white” and “not a van” for vehicles and “not red” and “not a circle” for the abstract task for DA and MT inferences respectively) and contrast class (using “blue” for vehicles and “yellow” for the abstract tasks in DA inferences and “a motorbike” for vehicles and “a square” for the abstract task for MT inferences) conditions.

While the previous two experiments provided learning phases which were intended to affect responses to MP and AC inferences, the current experiment will have two learning phase conditions for each scenario intended to affect either DA or MT inferences. This is based on the further development of the approach proposed by Oaksford and Chater (2007). Each participant will see one scenario intended to suggest that DA inferences are more likely when “not” negation is used in place of the contrast class and no difference in MT inferences between these conditions. The other scenario will keep the likelihood of DA inferences constant between “not” negated and contrast class conditions while suggesting that MT is more likely where “not” negation is used than the contrast class.

Table 4.7 – Contextual information provided – DA manipulation

		Colour of vehicle / Colour of shape			Total
		White vehicle / Red shape	Blue vehicle / Yellow shape	Red vehicle / Blue shape	
Type of vehicle /	Van / Circle	8	2	0	10
Type of shape	Motorbike / Square	0	0	10	10
	Car / Triangle	0	0	10	10
Total		8	2	20	30

Number of given type and colour presented in learning phase

In the condition where DA is varied, the number of vehicles or shapes of different types and colours presented in the learning phase (see table 4.7) suggests that the probability of a DA inference being correct when “not” negation is used in the minor premise is 91%. However, when a contrast class member is used in the minor premise, this probability drops to 0%. Given this context, it is expected that participants make fewer DA inferences when the contrast class member is used in the minor premise instead of the “not” negated premise. For the same reason, confidence ratings are expected to be lower when drawing the DA inference using the contrast class member in the minor premise than when explicit negation is used.

Where MT inferences are tested in this scenario, no significant differences are anticipated between “not” negated and contrast class conditions in responses to the inference task questions. Based on the learning phase materials in the condition, the probability that MT is correct is the same whether “not” negation is used (100%) or whether the contrast class is used (100%).

Table 4.8 – Contextual information provided – MT manipulation

		Colour of vehicle / Colour of shape			Total
		White vehicle / Red shape	Blue vehicle / Yellow shape	Red vehicle / Blue shape	
Type of vehicle /	Van / Circle	8	0	0	8
Type of shape	Motorbike / Square	2	0	0	2
	Car / Triangle	0	10	10	20
Total		10	10	10	30

Number of given type and colour presented in learning phase

In the MT manipulation condition, given the conditional ‘*if it is white, then it is a van*’, the number of vehicles or shapes of different types and colours presented in the learning phase (see table 4.8) suggests that the probability of a MT inference being correct when ‘*it is not a van*’ is used in the minor premise is 91%. However, when the ‘*motorbike*’ contrast class member is used as the minor premise, this probability drops to 0%. Therefore, it is expected that participants make fewer MT inferences when the contrast class member is used in the minor premise instead of the “not” negated premise in this condition. Confidence ratings are expected to be lower when drawing the MT inference using the contrast class member in the minor premise than when explicit negation is used in this condition.

No significant differences are anticipated in responses to DA inferences in this condition. The learning phases in this condition do not suggest any difference in the probability that the DA inference is correct if “not” negation is used (100%) or if the contrast class is used (100%).

#### 4.4.2. Method

##### 4.4.2.1. Design

Like the previous experiment, a mixed design was used with inference type (MP, DA, AC and MT), conclusion polarity (“not” negated and contrast class for DA and MT inference types and affirmative for MP and AC types) and manipulation (DA manipulation and MT manipulation) as within-participant factors. As with the prior experiment, scenario was a between participants factor. All participants saw both

scenarios, but half saw the vehicle scenario in relation to the MP manipulation condition questions and the other half seeing it in relation to the AC manipulation conditions (with the other scenario used for the alternative manipulation condition).

Dependent variables were, again, responses to the inference tasks (whether the inference was made or not) and confidence in that response (based on a visual analogue scale using a slider providing a response between 1 and 100).

#### 4.4.2.2. Participants

Participants were recruited through the ‘Amazon Mechanical Turk’ system and received a small payment (US\$0.50) after completion. A small number of respondents were excluded because their uniqueness could not be confirmed (matching either Mechanical Turk ID or IP address) leaving a total of 168 unique participants. The participants had median age 34 (range 19 to 75), were 56% female and 96% had English as their first language.

#### 4.4.2.3. Materials

The scenario involving vehicles used in the previous experiment was used again in the present experiment (see §4.3.2.3). Again, two sets of 30 learning phase tasks were prepared with images of vehicles of different types and colours. However, this time the number of learning tasks of each type and colour matched those required for the DA variation and MT variation conditions which are set out above. The inference tasks with the vehicle scenario used a different inference rule: ‘*If it is white, then it is a van*’. The minor premises used in each inference task were also changed to the following:

- *white*. [MP]
- *not white*. [DA with “not” negation]
- *blue*. [DA with contrast class]
- *a van*. [AC]
- *not a van*. [MT with “not” negation]
- *a motorbike*. [MT with contrast class]

The options available for responses to the vehicle inference questions were the same as the previous experiment.

A new scenario using more abstract materials was also developed which involved a quality control manager checking cards with shapes of different types and colours that a machine produced. Two sets of 30 learning phase tasks showing photos of cards with shapes of different types and colours in line with the frequencies required for the DA variation and MT variation conditions were prepared. Photos showed cards on a white background and were cropped and re-sized to ensure each learning task had the same sized picture. Each learning phase task in this scenario asked the following questions:

*What type of shape is this?* [Answer options: *Circle; Square; Triangle*]

*What colour is this shape?* [Answer options: *Red; Yellow; Blue*]

Six inference tasks were generated for this scenario as follows:

*The quality control manager is considering the following rule about the cards that she sees:*

- *If it is red, then it is a circle.*

*A system on the machine indicates that the shape on the next card is:*

[One of the following minor premises was provided for each question]

- *red.* [MP]
- *not red.* [DA with “not” negation]
- *yellow.* [DA with contrast class]
- *a circle.* [AC]
- *not a circle.* [MT with “not” negation]
- *a square.* [MT with contrast class]

*Please select the option below that best describes what she should conclude about the next shape.:*

[One of the following pairs of options was provided with the options in random order for each question]

[For MP and DA questions:]

*That the shape is a circle*

*That the shape is not a circle*

[For AC and MT questions:]

*That the shape is red*

*That the shape is not red*

Full details of the scenarios used in this experiment and questions asked are included in appendix 3 (see §A3.2 for the ‘vehicles’ scenario and §A3.3 for the ‘shapes’ scenario).

#### 4.4.2.4. Procedure

After being recruited through the Amazon Mechanical Turk System ([www.mturk.com](http://www.mturk.com)), participants were directed to a web-based survey (using [www.surveymzmo.com](http://www.surveymzmo.com)) to complete the experiment. The survey started with an information screen requiring participants to confirm consent to progress. Participants were randomly allocated to see the DA variation condition using the vehicle scenario and MT variation condition using the abstract (shapes) scenarios or vice versa. They were also randomly allocated to seeing the vehicle scenario first or second. Participants were asked to provide basic demographic information and then overall instructions and instructions for the first learning phase were provided.

The learning phases each had thirty randomly ordered pages. A learning phase page consisted of a picture of either a vehicle or a card with a shape on it depending on the scenario and two questions. The first question asked what type of vehicle or shape was shown (with answer options *car*, *van* and *motorbike* for the vehicle scenario and *circle*, *square* and *triangle* for the abstract scenario presented in random order). The second question asked what colour the vehicle or shape was (with available responses *white*, *blue* and *red* provided in random order for vehicles and *blue*, *yellow* and *red* provided in random order for the shapes scenario). Participants had to answer both questions on each page before they could proceed.

Instructions for the inference questions followed each learning phase. The six inference questions for each scenario (as described in §4.4.2.3) were each presented on a single page in random order. On each page, participants were asked to pick an option to respond to the inference question and click on a slider bar to indicate their confidence



that their response was correct. Responses to the confidence question were recorded at a number from 1 to 100 based on where they clicked on the bar (with 1 representing the end labelled ‘Not at all confident’ and 100 the end labelled ‘Completely confident’.)

A verification task on a single phase followed each scenario’s inference question. This task asked participants to indicate what number of the next 100 vehicles or shapes they would expect to fall into each type and colour category. They were required to respond in a three by three grid. The columns were labelled in random order with the possible colours used in the scenario (*white, blue, red* for vehicles and *red, yellow, blue* for the abstract task). The rows were labelled in random order with types (*car, van, motorbike* for vehicles and *circle, square, triangle* for the abstract tasks). Participants’ responses were required to sum to 100 before they could proceed.

Participants completed the learning, inference and verification tasks of the second scenario after they had completed the first.

Finally, participants were thanked for their time and provided with a code to use on the Mechanical Turk system which would confirm that they had completed the survey. They were also provided with contact information if they had any questions.

#### **4.4.3. Results**

Scores reflecting participants’ confidence in each inference considered were derived from responses to the inference and confidence questions, as with the prior two experiments. The proportion of responses supporting the relevant inference and average confidence scores are set out in table 4.9.

Table 4.9 – Summary of responses to chapter 4, experiment 3

	Trials in DA manipulation condition					
	MP	DA		AC	MT	
		Not	Contrast		Not	Contrast
Proportion making Inference	94.0%	94.0%	55.4%	95.2%	95.8%	94.6%
Confidence of Inference (standard deviation)	.84 (0.23)	.72 (0.27)	.48 (0.35)	.76 (0.24)	.78 (0.27)	.80 (0.26)
	Trials in MT manipulation condition					
	MP	DA		AC	MT	
		Not	Contrast		Not	Contrast
Proportion making Inference	95.2%	92.3%	95.8%	95.2%	94.6%	61.9%
Confidence of Inference (standard deviation)	.78 (0.23)	.77 (0.28)	.79 (0.28)	.82 (0.23)	.68 (0.27)	.47 (0.35)
Proportion making inference as a percentage to 2 decimal places						
Mean Confidence (standard deviation) to 2 decimal places						

Planned comparisons using McNemar  $\chi^2$  tests were used to compare the number of responses supporting the inference in the “not” negated and contrast class conditions for the DA and MT inferences in each manipulation condition. In the DA manipulation condition, these found a significant difference between the number of inferences supported in “not” negated and contrast class conditions for DA inferences ( $\chi^2(1) = 63.060, p < .001$ ) but not MT inferences. In the MT manipulation, no difference was found in the number of DA inferences endorsed between the “not” negated and contrast class conditions, but a significant difference was found in the number of MT inferences endorsed between the two negation conditions ( $\chi^2(1) = 49.590, p < .001$ ).

Confidence scores were compared using a 2x2x2x2 mixed measures ANOVA. Manipulation (DA manipulated and MT manipulated), inference (DA and MT) and negation used (“not” negation and contrast class) were the repeated measures factors. Scenario (the vehicles scenario in the DA manipulation and shapes scenario in the MT

manipulation and vice versa) was the independent measures factor. A significant three-way interaction between manipulation, inference and negation type ( $F(1,166) = 129.132, p < .001, \eta_p^2 = .438$ ), a significant two way interaction between manipulation and inference ( $F(1,166) = 97.411, p < .001, \eta_p^2 = .370$ ) and a significant main effect of negation ( $F(1,166) = 101.426, p < .001, \eta_p^2 = .379$ ) were found. No main effect of scenario was found not any interactions with this factor. For this reason, material effects are not examined further in this experiment.

To understand the interaction between negation and inference types, four planned comparison t-tests were performed (with Bonferroni adjusted significance levels of  $p = .013$ ). In the DA manipulation condition, there was a significant difference in confidence scores depending on which negation type was used for DA inferences ( $t(167) = 7.782, p < .001$ ) but not MT inferences. In the MT manipulation condition there was a significant difference in confidence scores depending on which negation type was used for MT inferences ( $t(167) = 7.151, p < .001$ ) but not DA inferences.

#### 4.4.3.1. Engagement with Learning Phase

Across both learning phases, each participant was asked 120 questions about the pictures that they saw. About 42% of respondents made any mistakes at all on the learning task questions (71 participants). The number of mistakes made by each participant across both learning tasks ranged between 0 and 40, with a median of 0, a mean of 1.1 and a standard deviation of 3.67.

At the end of each scenario, participants were asked for their own estimates indicating the proportion of vehicles or shapes of each type and colour.  $R^2$  statistics were then calculated for each participant in each scenario which indicated how well their estimates approximated information provided in the learning stage (the  $R^2$  statistic is described in §4.2.4.1). A range from  $R^2 = -2.125$  to  $R^2 = 1.000$  was observed (median  $R^2$  was 0.656, mean  $R^2$  was -35.794 and the standard deviation of  $R^2$  scores was 255.90). To look only at responses from those participants whose internal model of the scenario appeared to reflect, to at least some extent, the information presented in the learning phase, responses to scenarios where the  $R^2$  score was less than zero were removed from the data. This left 122 responses to each inference and confidence question in the DA manipulation condition and 116 responses to each question in the MT manipulation

condition (both reduced from 168 sets of responses in each condition). The analysis above was repeated with this sub-sample with manipulation condition treated as a between participants variable (because some participants responses in one condition were now excluded).

As with the full sample, McNemar  $\chi^2$  tests found differences between the number of inferences endorsed by the sub-sample depending on the negation type used for DA inferences in the DA manipulation condition ( $\chi^2 (1) = 58.065, p < .001$ ) and MT inferences in the MT manipulation condition ( $\chi^2 (1) = 49.076, p < .001$ ). No differences were found between responses where different negation types were used for MT inferences in the DA manipulation condition or DA inferences in the MT manipulation condition. With this smaller sample, inference levels for DA were 95.9% and 46.7% respectively when “not” negation and contrast classes were used in the DA manipulation condition and 96.6% and 97.4% respectively in the MT manipulation condition. Inference levels for MT were 98.4% when both “not” negation and contrast classes were used in the DA manipulation condition and 97.4% and 53.4% respectively in the MT manipulation condition.

A 2x2x2x2 mixed measures ANOVA was completed using this smaller sample with inference (DA and MT) and negation (“not” negated and contrast classes) types as within participants factors and inference manipulated (DA and MT) and scenario (vehicle scenario in DA manipulation and shapes scenario in MT manipulation and vice versa) as between participants factors. As with the similar ANOVA for the full sample, this found a significant three way interaction between manipulation, inference and negation type ( $F(1,234) = 160.857, p < .001, \eta_p^2 = .407$ ), a significant two way interaction between manipulation and inference ( $F(1,234) = 257.708, p < .001, \eta_p^2 = .524$ ) and a significant main effect of negation ( $F(1,234) = 95.437, p < .001, \eta_p^2 = .290$ ). In addition, interactions between inference type and scenario ( $F(1,234) = 4.320, p = .039, \eta_p^2 = .018$ ), between negation type and scenario ( $F(1,234) = 10.103, p = .002, \eta_p^2 = .041$ ), between inference, scenario and manipulation ( $F(1,234) = 6.559, p = .011, \eta_p^2 = .027$ ) and between inference, negation, scenario and manipulation ( $F(1,234) = 9.382, p = .002, \eta_p^2 = .039$ ) were found. All of these interactions involving scenario are small compared to those effects that were also observed in the full sample (and most are at the margins of significance).

The four planned comparison t-tests (with Bonferroni adjusted significance levels of  $p=.013$ ) compared confidence scores between the “not” negated and contrast class conditions for the DA and MT inferences in each of the DA manipulation and the MT manipulation conditions for this sub-sample. As with the full sample, confidence scores differed significantly based on the type of negation used for DA inferences in the DA manipulation condition ( $t(121) = 8.182, p<.001$ ) and MT inferences in the MT manipulation condition ( $t(115) = 8.997, p<.001$ ). Neither the comparison for MT inferences in the DA manipulation condition ( $t(121) = -0.086, p=.932$ ) nor for DA inferences in the MT manipulation condition ( $t(115) = -2.245, p=.027$ ) reached the required level of significance. The mean confidence levels for these inferences is shown in figure 4.3 (error bars represent standard error).

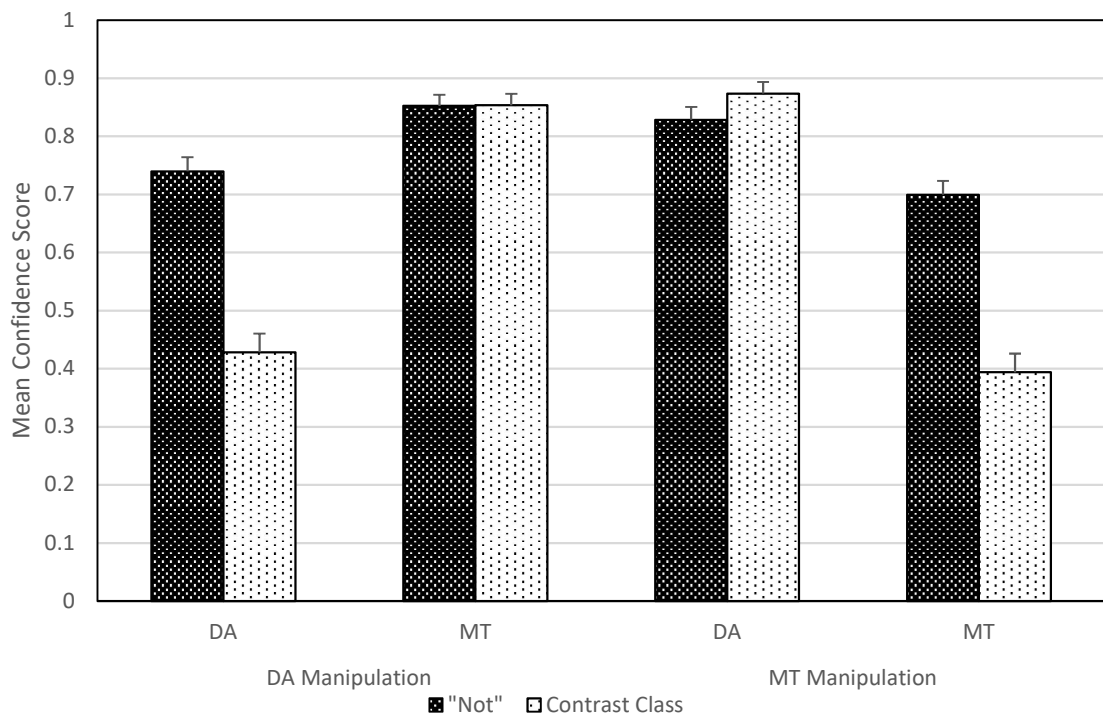


Figure 4.3 – Chart Showing Mean Confidence Scores for DA and MT Inferences of Participants that Demonstrated Learning in Experiment 3.

#### 4.4.4. Discussion

This experiment has extended the results seen with MP and AC in the previous experiment to DA and MT inferences. It has shown that frequency information can

create and remove an implicit negation effect in DA and MT inferences based on the conditional probabilities implied. There were no unexpected effects when only the subsample that showed evidence of learning frequency information were included.

Like the previous experiment, these results provide some substantial support for the fundamental insight of Oaksford et al.'s (2000) conditional probability model. An effect was created when the conditional probabilities implied that it should be and was absent where they implied that it should not be. This and the prior experiment have shown that this effect can be generated in MP, DA, AC and MT. The size of the effect in experiments 2 and 3 was much larger than in experiment 1 (where the difference in conditional probabilities motivating the effect was smaller).

These effects are also hard for other theories to explain. The fact that the effect can be induced and removed based on conditional probabilities, in particular, makes it hard to see how mental models theory (e.g. Johnson-Laird & Byrne, 2002) or heuristic-analytic theory without a probabilistic analytic stage (e.g. Evans, 1984) could account for the data.

Evidence of learning frequency information was similar to the previous experiment (70.8% of cases had  $R^2 > 0$  in this experiment which is close to the 70.5% seen previously). This experiment, like the last one, presented the learning task before any other frequency information is given or inference tasks completed. The high level of apparent learning supports the idea that participants engaged better where the learning phase is early in the experimental procedure.

This experiment also set out to test whether contrast class manipulation effects would remain where abstract rather than naturalistic stimuli were used. In the ANOVAs conducted, no main effect of group was observed which would have been anticipated if one scenario suppressed the anticipated effects. When a sub-group of those that demonstrated engagement with the learning task was considered, some interactions involving the group factor were observed. However, the size of these effects was tiny compared to the effect of the interactions between inference, negation and manipulation and between inference and manipulation interaction as well as the main effect of negation. For this reason, it appears that the abstract scenario behaved in the same

manner as the previously used naturalistic scenarios. This suggests that the conditional probability approach which formed the basis of the predictions for this experiment can be used to predict effects with abstract scenarios like those used by Evans and Handley (1999) when they first identified an implicit negation bias.

## **4.5. General Discussion**

This chapter set out to investigate Oaksford and Chater's (2007) intuitions about how implicit negation effects were created when contrast class members were used which were based on their conditional probability model. The results of all three experiments have been consistent with those intuitions, and the extension of those intuitions from MP to DA, AC and MT inferences.

Experiment 1 demonstrated the basic effect, that MP inferences could be suppressed where appropriate contrast class members were used to provide implicit negation of the premise. This creates an effect like that observed by Evans and Handley (1999) under the circumstances predicted by Oaksford and Chater (2007). However, it does not allow us to confirm that this effect is generated by the conditional probabilities implied by the frequency data provided.

Experiments 2 and 3 extended this result and demonstrated that the implicit conclusion suppression effect could be both created and removed just through the manipulation of conditional probabilities. Comparing the results of experiment 2 with experiment 1 show that the size of this effect can be increased by using more extreme differences in the conditional probabilities implied by the frequency information.

Experiment 3 also showed that this effect could be seen and manipulated with abstract as well as naturalistic materials.

Altogether, these experiments show firstly that the conditional probability approach to reasoning can provide an effective account of the implicit negation effect as seen by Evans and Handley (1999) and Schroyens et al. (2000). For the reasons discussed in §4.1.3, these experiments also create effects that are hard for mental models theory (e.g. Johnson-Laird & Byrne, 2002; Khemlani et al., 2012) or heuristic-analytic theory

without a probabilistic analytic stage (e.g. Evans, 1984) to account for (e.g. the creation and removal of the effect through manipulation of frequency data alone).

They have also shown that the same inferences can take different conditional probabilities in different circumstances, addressing Schroyens and Schaeken's (2003) criticism of the conditional probability approach proposed by Oaksford et al. (2000).

A final result of potential interest is that engagement with the learning phase appears to have been much stronger in experiments 2 and 3 than experiment 1. This may indicate that providing probability data up front or asking conditional inference questions before the learning phases (which was done in experiment 1 but not experiments 2 and 3) suppressed engagement in the learning phase. It is possible that people assumed they already had the information they needed from the initial information provided and therefore made less attempt to consider what they were seeing.



## **5. Negation in Reasoning – Conclusions and Next Steps**

This thesis has explored how the use of alternative negation – negation without “not” – can affect people’s responses to conditional reasoning problems that use negation. Two sets of experiments have considered how the use of alternative forms of negation (particularly affixal negation) affect people’s probability judgements of propositions with negation and how this impacts their reasoning performance. The third set of experiments has used implicit negation based on contrast classes to explore how frequency information about a negated concept can suppress conditional inferences.

In this chapter, I provide a summary of the experiments in this thesis and the key findings. I look again at how these findings might help us evaluate proposed accounts of reasoning. Finally, I consider some potential next steps to further evaluate the role of negation in human reasoning.

### **5.1. Key Findings**

This thesis presents three sets of experiments which aim to look at how considering different types of negation can inform our understanding of how people complete reasoning tasks. The first two sets of experiments in this study explored how using different forms of negation might affect deductive reasoning with conditionals. The first set looked at how different types of negation affected probability judgements. The second set looked directly at how different types of negation affected conditional inference judgements. The final set of experiments moved on to test predictions which depend on a contrast class account of negations comparing inference levels depending on whether “not” negation or implicit negation using a contrast class member was used.

#### ***5.1.1. Different Types of Negation and Probability***

The first set of experiments looked at how the use of different expressions of negation in propositions affected probability judgements in relation to those propositions. Different types of negation can lead to different understandings with more contrary or contradictory interpretations (e.g. Horn, 1989; Zimmer, 1964). A proposition that uses negation with a more contrary interpretation allows a narrower interpretation than the same proposition with a more contradictory form of negation (Horn, 1989). Bianchi, Savardi & Kubovy (2011) found that people can and do distinguish between the broader range of possibilities provided by the (contradictory) negation of concepts and the

contrary implication of the concept (the opposite). It was, therefore, anticipated that participants would rate the probability of a proposition using more contrary forms of negation lower than where more contradictory forms are used.

Prior research on conditional reasoning has typically treated negation as a simple operator. Most key experiments looking at negation effects have used either “not” negation with Evans’ (1977) conditional inference paradigm or implicit negation in Wason’s (1968) selection task. Consideration of the nature of negation has been rare. Experiments have shown that the probability of propositions used in conditionals can affect biases in conditional reasoning (e.g. Oaksford, Chater & Larkin, 2000). Therefore, if different types of negation gave rise to different probabilistic implications, then this might help us understand how negation effects lead to reasoning biases.

The first experiment looked at several different types of affixal negation (using affixes “iN-”, “un-” and “non-”), negation using “not” and implicit negation (based on opposites used in Bianchi et al.’s, 2011, materials). This found a clear difference between the understanding created by propositions using very highly contrary implicit negation and the other (less contrary) forms of negation tested. However, it did not find the anticipated differences between the other forms of negation which varied in their levels of contrary and contradictory interpretation.

The second experiment sought to look at whether more sensitive materials could find a difference between affixal and “not” negation. It, therefore, used materials allowing all forms of negation to be used with each proposition and just compared negation using affixes (“iN-”, “un-” and “non-”) and “not”. However, the design meant that an individual’s ratings of each type of negation could not be directly compared. The results again showed a difference between the form of negation that was expected to be most contrary (“iN-”) and the other forms of negation but none of the differences expected between those other forms of negation.

The final experiment in this first set used materials that better reflected how affixal negation is used in natural language. It aimed to see whether there was a difference between the probability ratings of propositions depending on whether “not” or “un-”

negation was used. The results only showed a significant effect when materials that may have been ambiguous were removed so should be treated cautiously.

This series of experiments has therefore shown that different types of negation can lead to different probability ratings when they are used in propositions. This suggests that using different types of negation in reasoning problems is likely to have an effect of the biases observed. Each experiment, in turn, was able to use more sensitive materials to show that differences between more similar forms of negation could be found. The size of effects observed was small and further experiments could potentially allow the nature of these effects to be better explored. However, this first set of experiments demonstrated that the use of different types of negation had the potential to effect reasoning biases so the next chapter would look directly at their use in conditional inference tasks.

### ***5.1.2. Different Types of Negation in Conditional Inference***

The second set of experiments looked directly at how different types of negation might affect conditional inferences. As the first set of experiments had shown, the use of different types of negation in propositions could alter the perceived probability of those propositions. Prior experiments often looked at the use of negations in conditional inference tasks using Evans' (1977) paradigm which looked at negation effects by considering negation as a single, simple operator. This set of experiments used a novel extension of this paradigm which allowed consideration of two forms of negation. These experiments aimed to help evaluate different models of reasoning. In particular, differential effects of different types of negation would suggest support for models that allowed support for models based on causality and probability (e.g. Fernback & Erb, 2013; Oaksford et al., 2000). Other models, such as those based on mental logic (e.g. Braine & O'Brien, 1991) or mental models (e.g. Johnson-Laird, Byrne & Schaeken, 1992) suggest reasoning depends on algorithms which only allow for negation to exist as a simple contradictory operator. Such models are unable to account for the differences anticipated in these experiments.

The first experiment in this set varied antecedent polarity in conditionals between three conditions (affirmative, "not" negated and "un-" negated). It was intended as an initial test of whether different forms of negation would lead to different biases in conditional

inference tasks. The results showed that inference levels for MP, DA and MT varied between polarity conditions in the directions predicted. However, responses using the alternative negation condition (“un-”) did not significantly differ from either the affirmative or “not” negated conditions. The different order seen in AC negation may be explainable as an artefact of the oft-observed negative conclusion bias (e.g. Evans, Clibbens & Rood, 1995). Overall these results are ambiguous about the predicted effect of alternative negation. However, a very significant effect was observed when high plausibility conditionals were used compared to lower plausibility conditionals. The next experiment would seek to control for this effect with the aim of focusing on any independent effect of an alternative form of negation.

The second experiment attempted to control for the plausibility effects observed in the first experiment. It did this while testing three polarity conditions (affirmative, “not” and “un-”) across both antecedent and consequent terms in conditionals. The results showed that MP inferences approached ceiling but, unusually, there were minimal differences in the levels of DA, AC and MT inferences endorsed. This may be the result of using low plausibility conditionals used in this experiment (to control for plausibility effects). Therefore, while the results of this experiment do not help us understand how different types of negation may have differential effects, they do provide substantial support for models (including causal and probabilistic models) which would suggest that inference levels are influenced by conditional plausibility.

The third and fourth experiments in this set used the same sets of materials. These materials were intended to test three polarity conditions (affirmative, “not” and “un-”) in both the antecedent and consequent terms of conditionals. The materials were intended to have normal plausibility characteristics (and therefore the plausibility of conditionals was also expected to affect responses). The third experiment used a design in which each respondent answered one question for each inference type (with randomised polarity conditions). The fourth experiment used a within-participants design with all participants answering questions using all inferences in all polarity conditions. Neither experiment showed different responses depending on the different types of negation used as had been anticipated. However, both provided results that supported probabilistic and causal models of reasoning. For example, there was variation in results between the fully within participants design in experiment 4 and the

experiment 3 design in which participants saw each inference once in random polarity conditions. The suppression of MP inferences in the experiment 3 design but not the within participants design of experiment 4 is consistent with participants generating more disablers when a scenario is considered for the first time under Fernbach and Erb's (2013) Causal Bayesian Networks model.

Overall these experiments did not find the robust effects of different types of negation which had been sought. However, they did provide additional evidence to support models of reasoning based on Causal Bayesian Networks (e.g. Fernbach & Erb, 2013) and probabilistic computational aims (e.g. Oaksford et al., 2000). There are some indications that there is an effect of different types of negation but any such effect is much smaller than the effect of the plausibility of conditionals. A further experiment could attempt to provide a more sensitive analysis of different types of negation in conditional reasoning. However, given the evidence that any such effects are easily overwhelmed by other factors, it seemed appropriate to look at a different form of alternative negation for the final set of experiments.

### **5.1.3. Contrast Classes in Conditional Inference**

The final set of experiments returned to considering the differences between “not” negation and implicit negation which had previously been considered in the first of the initial set of experiments. In particular, these experiments looked at how members of the contrast class (the set of concepts that could replace a negated concept; Oaksford & Stenning, 1992) could be used to provide implicit negation in a context in which probabilities were manipulated. These experiments specifically aimed to test and extend a hypothesis proposed by Oaksford and Chater (2007) that inference levels could be reduced in logically equivalent questions where the probabilities of propositions involving contrast classes are varied.

These experiments used a learning phase to provide probability information to participants. Similar learning phases have been used in relation to Wason's (1968) selection task (e.g. Pollard & Evans, 1982; Oaksford & Wakefield, 2003). However, such a learning phase has not previously been employed in conditional inference tasks and nor have they previously only involved simple questions about what is seen (rather than relating to the conditional being considered). In these experiments, the frequency

information about the occurrence of contrast class members was varied, so given contrast class members were either high probability or low probability. Oaksford and Chater (2007) predicted that when the contrast class members were used to provide implicit negation in a conditional inference task, the level of MP inferences could be varied depending on whether the contrast class member was high or low probability. Such a finding would support a probabilistic account of conditional reasoning (e.g. Oaksford et al., 2000).

The first experiment in this set directly tested Oaksford and Chater's (2007) hypothesis that given appropriate frequency information, the level of MP inferences will vary depending on whether "not" negation or implicit negation using a contrast class member was used in the inference task. The results showed substantial suppression of MP inferences when implicit negation was used. This provided some initial support for the hypothesis. However, this was not sufficient to confirm that the probability information was responsible for the suppression. The next experiment would, therefore, provide conditions that would show that inferences in the contrast class condition could be suppressed or not based on the probability information provided.

The second experiment provided two conditions in which different frequency information was provided. It also extended Oaksford and Chater's (2007) hypothesis to apply to AC inferences. One condition was intended to leave MP inferences at similar levels when "not" negation and an implicitly negating contrast class member were used but suppress AC inferences when implicit negation was used compared to "not" negation. The other condition reversed this, with MP inferences expected to vary but AC inferences expected to be stable across "not" negated and contrast class conditions. The results were in line with these predictions.

The final experiment sought to manipulate DA and MT inferences in the same manner as the prior experiment had manipulated MP and AC (extending Oaksford & Chater's, 2007, hypothesis to these inferences). It also included a condition which used abstract rather than naturalistic materials in the learning phase and experimental tasks. The results showed that DA and MT inference levels could also be manipulated by varying the probability of contrast class members. It also showed that these effects are consistent when abstract and naturalistic materials were used.

These results provide considerable support for probabilistic models of reasoning (e.g. Oaksford et al., 2000) and to the relevance of the contrast class account of the understanding of negation in reasoning (Oaksford & Stenning, 1992). The ability to manipulate inference levels using frequency information implying probability levels cannot be accounted for in current mental logic (e.g. Braine & O'Brien, 1991) or mental models (e.g. Johnson-Laird et al., 1992) approaches.

## **5.2. Accounts of Reasoning**

As set out in §1.3, a number of models have been developed to provide an account of reasoning biases. The experiments discussed above sought to look at whether the use of different types of negation in reasoning problems can help us evaluate these competing accounts. I therefore, briefly, revisit these accounts in this section to discuss how our understanding might have been developed in this thesis.

### **5.2.1. Mental Logics**

Approaches based on mental logics (e.g. Braine & O'Brian, 1991; Rips, 1994) essentially suggest that models of reasoning should aim to replicate classical logic. They propose models with algorithms that seek to imitate classical logic but with limitations in memory and processing capacity which lead to reasoning biases.

As discussed in §1.3.1, these approaches suggest that people use internal rules for solving inference tasks. These rules are intended to emulate those of classical logic. However, sometimes an inference task will require a rule that people do not have in their mental logic and it is only through more complex reasoning that the task can be solved. Proponents of the mental logics approach suggest that reasoning biases occur where people's internal logic is incomplete. For example, Braine and O'Brian's (1991) mental logic approach suggests people have straightforward access to a rule of MP but not MT which requires more complicated processing. In their most basic form, mental logic approaches can explain basic reasoning biases (such as the easier acceptance of MP over MT). But, they do not appear to be able to take account of content, context and negation effects. They have therefore been extended. For example, Noveck and O'Brien (1996) have sought to use Cheng and Holyoak's (1985) pragmatic reasoning schemas to explain biases involving negation and Rips (1994) suggests that additional rules account

for negation effects. These extensions essentially allow context to provide an additional set of rules and/or constraints in which the basic mental logic operates.

No firm conclusions about mental logic approaches can be drawn from the first set of experiments in this thesis. These experiments did not directly look at conditional reasoning (only at factors previously shown to be related to conditional reasoning).

The second set of experiments did look directly at conditional reasoning. As discussed above, there were suggestions in these results that different forms of negation may affect reasoning. Basic mental logics treat negation as a simple logical operator and are therefore unable to account for differences motivated by different forms of negation. Approaches such as Rips' (1994) PSYCOP model do not appear to allow for a differential treatment of different types of negation within the rules based approach. Similarly, Noveck and O'Brien's (1996) application of pragmatic reasoning schemas, would appear to provide additional constraints which may lead to different biases depending on the presentation of the problem. However, they do not account for different biases where the problems are essentially the same except for the negative forms used. It may be possible to further augment mental logic approaches with additional rules (for example, an 'un-x' to 'not-x' conversion that complicates – and therefore makes less likely – drawing inferences) but no current model predicts such a process. The results suggested by this sets of experiments therefore have the potential to undermine mental logic approaches. However, as these results were not seen consistently or robustly across the four experiments, any conclusions undermining mental logic approaches should be treated provisionally.

The final set of experiments again looked directly at conditional reasoning and how it was affected by probability manipulations. In the second and third experiment of this set, the same inference questions were answered differently depending on the context provided by frequency data. It is not clear how the algorithms proposed in mental logic approaches could explain the strong observation that probability manipulation can suppress inferences. The frequency information would not be sufficient to suggest an alternative pragmatic reasoning schema. Nor would any rules affecting the different treatment of 'not' and implicit negation explain the effect (as questions using both types of negation were answered in different ways depending on the frequency information



provided. This set of experiments, therefore, provides evidence which appears inconsistent with the role of mental logic in conditional reasoning.

### **5.2.2. Mental Models**

The mental models approach (Johnson-Laird, Byrne & Schaeken, 1992) proposes that people consider models when reasoning and each model representing a possibility or set of possibilities (see §1.3.2). Reasoning biases are said to occur because generating models requires time and mental capacity so people often reason over an incomplete set of models. Khemlani, Orenes and Johnson-Laird (2012) seek to provide an account of negation in mental models theory which does consider the complex forms that negation can take and range of implications that it can create (see §4.1.3). However, this account is still limited to negation creating context effects that make reasoning tasks more or less complicated under certain circumstances.

Because the first set of experiments does not directly consider the function of conditional reasoning, firm conclusions about the mental models' approach cannot be drawn. Where the second set of experiments provide limited evidence that different forms of negation may affect reasoning, this may help us to evaluate the mental models approach. The basic approach (e.g. Johnson-Laird & Byrne, 2002) would not appear to allow a mechanism that provides for different levels of inference to be drawn when different types of negation are used (the models considered are essentially equivalent for each type of negation so would not predict different biases). However, Khemlani et al.'s (2012) proposals may provide an approach that allows for such results. They suggest that a general context effect applies in which more processing time is required where some different types of negation are used. This may help account for the data from these experiments. But, as before, any conclusions are limited at this stage as the results of the experiments in this set were inconsistent.

The final set of experiments did provide clear evidence that frequency manipulation can suppress inferences even where lexical information is constant. The inference tasks used between the inference manipulation conditions are identical (i.e. between the two MP manipulation conditions and the two AC manipulation conditions in experiment 2 and between the two DA manipulation conditions and the two MT manipulation conditions in experiment 3). Therefore, the same mental models should be generated in each

condition. The issue here cannot be additional complexity in processing the different type of negation alone (as Khemlani et al., 2012, suggests). Given the same mental models should be generated, and the lexical information in the inferences tasks does not vary (which might permit a difference in the effect proposed by Khemlani, 2012), it is not clear how the mental models approach can account for these results. Therefore, the results of these experiments would not appear to be consistent with the theory that people reason over mental models.

### **5.2.3. Probabilistic Approach**

Probabilistic accounts of reasoning propose that the computational aim of reasoning is to reflect probability theory rather than classical logic, as this is more relevant to everyday life (Oaksford & Chater, 2001). Models based on this approach have provided a good account of the biases seen in conditional reasoning (e.g. Oaksford et al., 2000). More recently, algorithmic models have been developed based on Causal Bayesian Networks which are consistent with probabilistic approaches (e.g. Fernbach and Erb, 2013).

The first set of experiments in this thesis considered people's probability ratings of propositions. Such differences in the perception of propositions are relevant to probabilistic approaches but not to other approaches. Therefore, the results showing that people do rate the probability of propositions differently depending on the type of negation used provides some support for the role of probabilistic approaches. In particular, it seems sensible that the computational aim of a model should be able to incorporate the different information that people consider – including probability information.

The second set of experiments also provided support for the probabilistic approach. This set of experiments suggested that different forms of negation may affect reasoning. Such results are easier to account for with probabilistic models than models that do not allow for different probabilistic interpretations of negation. However, these results are tentative as the expected differences were not seen consistently across the four experiments.

The final set of experiments showed that conditional reasoning could be affected by probability manipulations. The predictions confirmed in this experiment were based on the probabilistic approach (Oaksford & Chater, 2007). These findings, therefore, provide substantial support for the relevance of that approach in accounting for human reasoning.

#### **5.2.4. Multiple Processes**

Multiple process theories typically involve two or more systems which interact to process reasoning tasks. Typically, these theories combine an associative, automatic and quick system and an analytic, controlled and slow system (Stanovich & West, 2000). The former system places little burden on cognitive resources while the latter requires a high level of cognitive resources. Evans' (1984, 2006) Heuristic-Analytic Theory is a prominent multiple process model explored in some reasoning studies.

The current set of experiments neither support nor rule-out the involvement of multiple processes in principle. However, they do undermine specific approaches. For example, it is not clear how Evans' (1984) original theory could account for the impact of probability data in reasoning responses in the third set of experiments when the lexical materials in the tasks remained constant. Such results would suggest that any valid multiple process theory should have a substantial probabilistic element as part of one or more of its processes.

#### **5.2.5. Alternative Models**

In §1.3.5 a couple of other models were discussed: Cheng and Holyoak's (1985) account based on the use of pragmatic reasoning schemas and Politzer and Bonnefon's (2010) proposal of a calculus of possibilities.

As discussed, Cheng and Holyoak's (1985) approach ultimately seeks to distil the processing of reasoning problems into the implementation of logical rules. This aligns it with other approaches based on internal mental logics and it has been integrated with Braine and O'Brien's (1991) approach (e.g. Noveck & O'Brien, 1996). As discussed above, the general theme of the present experiments (that negation is a complex operator can act through probability manipulation) undermines support for such approaches.

Politzer and Bonnefon (2010) proposed that a calculus of possibilities may better account for observed reasoning biases than approaches based on standard propositional logic or probability. The current study does provide some support for this approach. In particular, the experiments in chapter 3 show that the plausibility of a conditional is strongly correlated with people's willingness to draw inferences using it. However, further consideration is required to establish the predictions that such an approach would make and how it could be distinguished from a probabilistic approach. This may be the subject of future research.

### **5.3. Next Steps**

While the current thesis has helped develop our understanding of the role of negation in reasoning, some of the findings – particularly based on the second set of experiments – are tentative. This final section, therefore, considers how these results could be built on and developed further in future studies.

The first set of experiments did show that people rate the probability of propositions using different types of negation differently. However, the effects found were small and easily influenced by other material effects. A further study seeking to demonstrate more robust and consistent results might want to use a within-participants design to look at people's responses where materials are directly comparable across all negation conditions (which was not the case in the first experiment in this set which did use a within-participants design).

The results in the second set of experiments were less conclusive. There were indications that different types of negation did affect people's response to reasoning problems. However, these effects appear to be obscured by larger scenario effects (particularly related to the plausibility of conditionals). A future experiment to develop these results should, therefore, consider using materials which controlled for plausibility effects (but did not have the artificial characteristics observed in experiment 2). It should use sufficient sets of materials to secure a large number of responses in a between-participants design (given the tendency seen in experiment 4 for people to make superadditive judgements when they have sight of all conditions).

The final set of experiments provided evidence for the role of the contrast class interpretation of negation in conditional reasoning. Further evidence for the use of contrast classes in reasoning could be done through direct checking for contrast class member priming when a given proposition is negated in a reasoning task. This could use lexical decision tasks (Meyer & Schvaneveldt, 1971) to explore priming effects in reasoning tasks.

## References

- Adams, E. W. (1987). On the meaning of the conditional. *Philosophical Topics*, 15, 5-22.
- Ahn, W., & Kalish, C. W. (2000). The role of mechanism beliefs in causal reasoning. In F. C. Keil, & R. A. Wilson (Eds.), *Explanation and Cognition*. Cambridge, MA: MIT Press.
- Ahn, W., Kalish, C. W., Medin, D. L., & Gelman, S. A. (1995). The role of covariation versus mechanism information in causal attribution. *Cognition*, 54, 299-352.
- Ali, N., Chater, N., & Oaksford, M. (2011). The mental representation of causal conditional reasoning: Mental models or causal models. *Cognition*, 119, 403-418.
- Aristotle (1923). *Metaphysics*. In W. D. Ross (Ed. & Trans), *The Works of Aristotle*. London, UK: Methuen & Co.
- Bianchi, I., Savardi, U., & Kubovy, M. (2011). Dimensions and their poles: A metric and topological approach to opposites. *Language and Cognitive Processes*, 26, 1232-1265.
- Braine, M. D. S., & O'Brien, D. P. (1991). A theory of if: A lexical entry, reasoning program, and pragmatic principles. *Psychological Review*, 98, 182-203.
- Buhrmester, M., Kwang T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, 6, 3-5.
- Chan, D., & Chau, F. (1994). Suppression of valid inferences: syntactic views, mental models, and relative salience. *Cognition*, 53, 217-238.
- Cheng, P. W., & Holyoak, K. J. (1985). Pragmatic reasoning schemas. *Cognitive Psychology*, 17, 391-416.
- Clark, H. H. (1974). Semantics and comprehension. *Current Trends in Linguistics*, 12, 1291-1428.
- Cosmides, L., & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgement under uncertainty. *Cognition*, 58, 1-73.

- Cummins, D. D. (1995). Naive theories and causal deduction. *Memory & Cognition*, 23, 646-658.
- Cummins, D. D., Lubart, T., Alksnis, O., & Rist, R. (1991). Conditional reasoning and causation. *Memory & Cognition*, 19, 274-282.
- De Neys, W., Schaeken, W., & D'Ydewalle, G. (2002). Causal conditional reasoning and semantic memory retrieval: A test of the semantic memory framework. *Memory and Cognition*, 30, 908-920.
- Edgington, D. (1995). On conditionals. *Mind*, 104, 235-329.
- Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. *American Psychologist*, 49, 709-724.
- Englebrechtsen, G. (1976). The square of opposition. *Notre Dame Journal of Formal Logic*, 17, 531-541.
- Evans, J. St B. T. (1972a). Interpretation and matching bias in a reasoning task. *Quarterly Journal of Experimental Psychology*, 24, 193-199.
- Evans, J. St B. T. (1972b). Reasoning with negatives. *British Journal of Psychology*, 63, 213-219.
- Evans, J. St B. T. (1977). Linguistic factors in reasoning. *Quarterly Journal of Experimental Psychology*, 29, 297-306.
- Evans, J. St B. T. (1983). Linguistic determinants of bias in conditional reasoning. *Quarterly Journal of Experimental Psychology*, 35A, 635-644.
- Evans, J. St B. T. (1984). Heuristic and analytic processes in reasoning. *British Journal of Psychology*, 75, 451-468.
- Evans, J. St B. T. (1993). The mental model theory of conditional reasoning: critical appraisal and revision. *Cognition*, 48, 1-20.
- Evans, J. St B. T. (1996). Deciding before you think: Relevant and reasoning in the selection task. *British Journal of Psychology*, 87, 223-240.
- Evans, J. St B. T. (1998). Matching bias in conditional reasoning: Do we understand it after 25 years? *Thinking and Reasoning*, 4, 45-82.
- Evans, J. St B. T. (2006). The heuristic-analytic theory of reasoning: Extension and evaluation. *Psychonomic Bulletin & Review*, 13, 378-395.

- Evans, J. St B. T., Clibbens, J., & Rood, B. (1995). Bias in conditional inference: Implications for mental models and mental logic. *Quarterly Journal of Experimental Psychology*, 48A, 644-670.
- Evans, J. St B. T., Clibbens, J., & Rood, B. (1996). The role of implicit and explicit negation in conditional reasoning bias. *Journal of Memory and Language*, 35, 392-409.
- Evans, J. St B. T., & Handley, S. J. (1999). The role of negation in conditional inference. *Quarterly Journal of Experimental Psychology*, 52A, 739-769.
- Evans, J. St B. T., Handley, S. J., & Over, D. E. (2003). Conditionals and conditional probability. *Journal of Experimental Psychology: Learning Memory and Cognition*, 29, 321-335.
- Evans, J. St B. T., & Lynch, J. S. (1973). Matching bias in the selection task. *British Journal of Psychology*, 64, 391-397.
- Evans, J. St B. T., & Newstead, S. E. (1980). A study of disjunctive reasoning. *Psychological Research*, 41, 373-388.
- Fernbach, P. M., & Erb, C. D. (2013). A quantitative causal model theory of conditional reasoning. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 39, 1327-1343.
- Gale, M., & Ball, L. J. (2006). Dual-goal facilitation in Wason's 2-4-6 task: What mediates successful rule discovery? *The Quarterly Journal of Experimental Psychology*, 59, 873-885.
- Gale, M., & Ball, L. J. (2012). Contrast class cues and performance facilitation in a hypothesis-testing task: evidence for an iterative counterfactual model. *Memory & Cognition*, 40, 408-419.
- Gigerenzer, G., Hoffrage, U., & Kleinbölting (1991). Probabilistic mental models: A Brunswickian theory of confidence. *Psychological Review*, 98, 506-528.
- Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, 102, 684-704.



- Giora, R., Fein, O., Metuki, N., & Stern, P. (2010). Negation as a metaphor-inducing operator. In L. R. Horn (Ed.), *The Expression of Negation*. Berlin, Germany: De Gruyter Mouton.
- Grant, S., Aithison, T., Henderson, E., Christie, J., Zare, S., McMurray, J., & Dargie, H. (1999). A comparison of the reproducibility and the sensitivity of visual analogue scales, Borg scales and Likert scales in normal subjects during submaximal exercise. *CHEST*, 116, 1208-1217.
- Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and Semantics 3: Speech Acts*. San Diego, CA: Academic Press.
- Griffen, D., & Tversky, A. (1992). The weighing of evidence and the determinants of confidence. *Cognitive Psychology*, 24, 411-435.
- Hasson, U., & Glucksberg, S. (2006). Does understanding negation entail affirmation? An examination of negated metaphors. *Journal of Pragmatics*, 38, 1015-1032.
- Hilton, D. J. (1995). The social context of reasoning: Conversational inference and rational judgement. *Psychological Bulletin*, 118, 248-271.
- Horn, L. R. (1985). Metalinguistic negation and pragmatic ambiguity. *Language*, 61, 121-174.
- Horn, L. R. (1989). *A Natural History of Negation*. Chicago, IL: University of Chicago Press.
- Jackson, S. L., & Griggs, R. A. (1990). The elusive pragmatic reasoning schemas effect. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, 42, 353-373.
- Jespersen, O. (1917). *Negation in English and Other Languages*. Copenhagen, Denmark: A. F. Høst.
- Johnson-Laird, P. N. (2008). Mental models and deductive reasoning. In J. E. Adler, & L. J. Rips (Eds), *Reasoning: Studies of Human Inference and its Foundations*. Cambridge, UK: Cambridge University Press.
- Johnson-Laird, P. N., & Byrne, R. M. J (1993). Precise of Deduction. *Behavioral and Brain Sciences*, 16, 323-380.

- Johnson-Laird, P. N., & Byrne, R. M. J (2002). Conditionals: A theory of meaning, pragmatics, and inference. *Psychological Review*, 109, 646-678.
- Johnson-Laird, P. N., Byrne, R. M. J., & Schaeken, W. (1992). Propositional reasoning by model. *Psychological Review*, 99, 418-439.
- Just, M. A., & Carpenter, P. A. (1971). Comprehension of negation with quantification. *Journal of Verbal Learning and Verbal Behaviour*, 10, 244-253.
- Karttunen, L., & Peters, S. (1979). Conventional Implicature. In C. Oh, & D. Dineen (Eds.), *Syntax and Semantics 11: Presupposition*. New York, NY: Academic Press.
- Kern, L. H., Mirels, H. L., & Hinshaw, V. G. (1983). Scientists' understanding of propositional logic: An experimental investigation. *Social Studies in Science*, 13, 131-146.
- Khemlani, S., Orenes, I., & Johnson-Laird, P. N. (2012). Negation: A theory of its meaning, representation, and use. *Journal of Cognitive Psychology*, 24, 541-559.
- Kirk, R. E. (1995). *Experimental Design – Procedures for the Behavioural Sciences*. London, UK: Brooks/Cole.
- Klima, E. (1964). Negation in English. In J. A. Fodor, & J. Katz (Eds), *The Structure of Language*. Englewood Cliffs, NJ.: Prentice Hall.
- Macchi, L., Osheron, D., & Krantz, D. H. (1999). A note on superadditive probability judgement. *Psychological Review*, 106, 210-214.
- MacDonald, M. C., & Just, M. A. (1989). Changes in activation levels with negation. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 15, 633-642.
- Manktelow, K. (2012). *Thinking and Reasoning: An Introduction to the Psychology of Reason, Judgement and Decision Making*. Hove, UK: Psychology Press.
- Manktelow, K. I., & Evans, J. St B. T. (1979). Facilitation of reasoning by realism: Effect or “non-”effect? *British Journal of Psychology*, 70, 477-488.
- Marcus, S. L., & Rips, L. J. (1979). Conditional Reasoning. *Journal of Verbal Learning and Verbal Behaviour*, 18, 199-223.

- Markovits, H. (1988). Conditional reasoning, representation, and empirical evidence on a concrete task. *The Quarterly Journal of Experimental Psychology*, 40A, 483-495.
- Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. San Francisco, CA: WH Freeman.
- Mazzon, G. (2004). *A History of English Negation*. Harlow, UK: Pearson Longman.
- Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, 90, 227-234.
- Noveck, I. A., & O'Brien, D. P. (1996). To what extent do pragmatic reasoning schemas affect performance on Wason's selection task? *Quarterly Journal of Experimental Psychology*, 49A, 463-489.
- Oaksford, M. (2002). Contrast classes and matching bias as explanations of the effects of negation on conditional reasoning. *Thinking and Reasoning*, 8, 135-151.
- Oaksford, M., & Chater, N. (1991). Against logicist cognitive science. *Mind and Language*, 6, 1-38.
- Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review*, 101, 608-631.
- Oaksford, M., & Chater, N. (1995). Theories of reasoning and the computational explanation of everyday inference. *Thinking & Reasoning*, 1, 121-152.
- Oaksford, M., & Chater, N. (2001). The probabilistic approach to human reasoning. *TRENDS in Cognitive Sciences*, 5, 349-357.
- Oaksford, M., & Chater, N. (2003). Computational levels and conditional inference: Reply to Schroyens and Schaeken (2003). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 150-156.
- Oaksford, M., & Chater, N. (2007). *Bayesian Rationality*. Oxford, UK: Oxford University Press.
- Oaksford, M., & Chater, N. (2010). Cognition and conditionals: An introduction. In M. Oaksford, & N. Chater (Eds.), *Cognition and Conditionals: Probability and Logic in Human Thinking*. Oxford, UK: Oxford University Press.

- Oaksford, M., & Chater, N. (2013). Dynamic inference and everyday conditional reasoning in the new paradigm. *Thinking & Reasoning*, 19, 346-379.
- Oaksford, M., Chater, N., & Grainger, B. (1999). Probabilistic effects in data selection. *Thinking & Reasoning*, 5, 193-243.
- Oaksford, M., Chater, N., & Larkin, J. (2000). Probabilities and polarity biases in conditional inference. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 26, 883-899.
- Oaksford, M., & Moussakowski, M. (2004). Negations and natural sampling in data selection: Ecological versus heuristic explanations of matching bias. *Memory & Cognition*, 32, 570-581.
- Oaksford, M., & Stenning, K. (1992). Reasoning with conditionals containing negated constituents. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 18, 835-854.
- Oaksford, M., & Wakefield, M. (2003). Data selection and natural sampling: Probabilities do matter. *Memory & Cognition*, 31, 143-154.
- O'Brien, D. P., & Manfrinati, A. (2010). The mental logic theory of conditional propositions. In M. Oaksford, & N. Chater (Eds), *Cognition and Conditionals: Probability and Logic in Human Thinking*. Oxford, UK: Oxford University Press.
- Oberauer, K., & Wilhelm, O. (2003). The meaning(s) of conditionals: Conditional probabilities, mental models and personal utilities. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 29, 680-693.
- Oberauer, K., Wilhelm, O., & Rosas Diaz, R. (1999). Bayesian rationality for the Wason selection task? A test of optimal data selection theory. *Thinking and Reasoning*, 5, 115-144.
- Ohm, E., & Thompson, V. A. (2006). Conditional probability and pragmatic conditionals: Dissociating truth and effectiveness. *Thinking and Reasoning*, 12, 257-280.
- Over, D. E., & Evans, J. St B. T. (2003). The probability of conditionals: The psychological evidence. *Mind and Language*, 18, 340-358.

- Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on Amazon Mechanical Turk. *Judgement and Decision Making*, 5, 411-419.
- Pearl, J. (2000). *Causality: Models, Reasoning and Inference*. Cambridge, UK: Cambridge University Press.
- Politzer, G., & Bonnefon, J. (2010). Two aspects of reasoning competence: A challenge for current accounts and a call for new conceptual tools. In M. Oaksford, & N. Chater (Eds.), *Cognition and Conditionals: Probability and Logic in Human Thinking*. Oxford, UK: Oxford University Press.
- Pollard, P., & Evans, J. St B. T. (1980). The influence of logic on conditional reasoning performance. *Quarterly Journal of Experimental Psychology*, 32, 605-624.
- Pollard, P., & Evans, J. St B. T. (1983). The effect of experimentally contrived experience on reasoning performance. *Psychological Research*, 45, 287-301.
- Popper, K. R. (1959). *The Logic of Scientific Discovery*. London, UK: Hutchinson & Co.
- Prado, J., & Noveck, I. (2006). How reaction time measures elucidate the matching bias and the way negations are processed. *Thinking and Reasoning*, 12, 309-328.
- Quinn, S., & Markovits, H. (1998). Conditional reasoning, causality, and the structure of semantic memory: strength of association as a predictive factor for content effects. *Cognition*, 68, B93-B101.
- Quinn, S., & Markovits, H. (2002). Conditional reasoning with causal premises: Evidence for a retrieval model. *Thinking and Reasoning*, 8, 179-191.
- Ramsey, F. P. (1931). *Foundations – Essays in Philosophy, Logic, Mathematics and Economics*. New Jersey: Humanities Press (1977).
- Rips, L. J. (1994). *The Psychology of Proof*. Cambridge, MA: MIT Press.
- Roberge, J. J. (1971). Some Effects of Negation on Adults Conditional Reasoning Abilities. *Psychological Reports*, 29, 839-844.
- Rosch, E. H. (1973). Natural categories. *Cognitive Psychology*, 4, 328-350.
- Rosch, E. H., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, 7, 573-605.

- Rumain, B. (1988). Syntactics of interpretation of negation: A developmental study. *Journal of Experimental Child Psychology*, 45, 119-140.
- Rumain, B., Connell, J., & Braine, M. D. S. (1983). Conversational comprehension processes are responsible for reasoning fallacies in children as well as adults: If is not the biconditional. *Developmental Psychology*, 19, 471-481.
- Russell, B. (1905). On denoting. *Mind*, 14, 479-493.
- Schroyens, W. J., & Schaeken, W. (2003). A critique of Oaksford, Chater, and Larkin's (2000) conditional probability model of conditional reasoning. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 29, 140-149.
- Schroyens, W. J., Schaeken, W., & d'Ydewalle, G. (2001). The processing of negations in conditional reasoning: A meta-analytic case study in mental model and/or mental logic theory. *Thinking and Reasoning*, 7, 121-172.
- Schroyens, W. J., Verschueren, N., Schaeken, W., & d'Ydewalle, G. (2000). Conditional reasoning with negations: Implicit and explicit affirmation or denial and the role of contrast classes. *Thinking and Reasoning*, 6, 221-251.
- Singmann, H., Klauer, K. C., & Beller, S. (2016). Probabilistic conditional reasoning: Disentangling form and content with the dual-source model. *Cognitive Psychology*, 88, 61-87.
- Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, 119, 3-22.
- Sloman, S. A. (2005). *Causal Models: How People Think About the World and its Alternatives*. Oxford, UK: Oxford University Press.
- Sommers, F. (1970). The calculus of terms. *Mind*, 79, 1-39.
- Sperber, D., Caram F., & Girotto, V. (1995). Relevance theory explains the selection task. *Cognition*, 57, 31-95.
- Stahl, C., Klauer, K. C., & Erdfelder, E. (2008). Matching bias in the selection task is not eliminated by explicit negations. *Thinking and Reasoning*, 14, 281-303.
- Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioural and Brain Sciences*, 23, 645-665.

- Stich, S. P. (1985). Could man be an irrational animal? Some notes on the epistemology of rationality, *Synthese*, 64, 115-135.
- Strawson, P. F. (1950). On referring. *Mind*, 59, 320-344.
- Taplin, J. E., & Staudenmayer, H. (1973). Interpretation of abstract conditional sentences in deductive reasoning. *Journal of Verbal Learning and Verbal Behaviour*, 12, 530-542.
- Thompson, V. A. (2010). Towards a metacognitive dual process theory of conditional reasoning. In M. Oaksford & N. Chater (Eds). *Cognition and Conditionals: Probability and Logic in Human Thinking*. Oxford, UK: Oxford University Press.
- Thompson, V. A., & Mann, J. M. (1995). Perceived necessity explains the dissociation between logic and meaning: The case of “Only If”. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 21, 1554-1567.
- Tottie, G. (1982). Where do negative sentences come from? *Studia Linguistica*, 36, 88-105.
- Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgement. *Psychological Review*, 90, 293-315.
- Vance, J. R. H. (2011). *Different Types of Negation in Reasoning*. Unpublished master’s dissertation, Birkbeck, University of London.
- Verschueren, N., & Schaeken, W. (2010). A multi-layered dual-process approach to conditional reasoning. In M. Oaksford & N. Chater (Eds), *Cognition and Conditionals: Probability and Logic in Human Thinking*. Oxford, UK: Oxford University Press.
- Von Wright, G. H. (1959). On the logic of negation. *Societas Scientiarum Fennica: Commentationes Physico-Mathematicae*, 22, 1-30.
- Wason, P. C. (1959). The processing of positive and negative information. *Quarterly Journal of Experimental Psychology*, 11, 92-107.
- Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology*, 12, 129-140.

- Wason, P.C. (1961). Response to affirmative and negative binary statements. *British Journal of Psychology*, 52, 133-142.
- Wason, P. C. (1965). The contexts of plausible denial. *Journal of Verbal Learning and Verbal Behavior*, 4, 7-11.
- Wason, P. C. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology*, 20, 273-281.
- White, P. A. (2000). Causal judgement from contingency information: The interpretation of factors common to all instances. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 26, 1083-1102.
- Wildman, T. M., & Fletcher, H. J. (1977). Developmental increases and decreases in solutions of conditional syllogism problems. *Developmental Psychology*, 13, 630-636.
- Yama, H. (2001). Matching versus optimal data selection in the Wason selection task. *Thinking and Reasoning*, 7, 295-311.
- Zimmer, K. E. (1964). *Affixal negation in English and other languages: An investigation of restricted productivity*. New York, NY: Linguistic Circle of New York.



## **List of Appendices**

Appendix 1. Materials Used in Chapter 2 Experiments

Appendix 2. Materials Used in Chapter 3 Experiments

Appendix 3. Materials Used in Chapter 4 Experiments

## **Appendix 1. Materials Used in Chapter 2 Experiments**

This appendix contains the experimental materials used in each of the three experiments reported in chapter 2. These considered how using different types of negation in propositions affected people's probability ratings of those propositions.

### **A1.1. Experiment 1**

The first experiment in chapter 2 looked at whether negation using affixes “non-”, “un-” and “iN-” and implicit negation led to different probability ratings than negation using “not”. This experiment is reported in §2.3. The method is set out in §2.3.2 and the development of the materials is discussed in §2.3.2.3.

#### **A1.1.1. Task Instructions**

The following text was provided to participants to explain the task:

*On each of the following pages, you will be given a scenario and asked to answer three questions about what you might expect given the scenario. Your response to each question will be an answer between 0 and 100.*

*This is asking for your view, there are no right or wrong answers and no questions are trying to catch you out.*

*You should try to complete each question as quickly as possible – go with your first reaction.*

*Once you've completed each question don't go back to it.*

#### **A1.1.2. Task Questions**

The following table sets out the 32 tasks presented to each participant. Each task provided a scenario and then three questions for the participants to answer. The tasks were presented in a booklet with the instructions (see above) on the front) and a task on each subsequent page. The order of the 32 tasks was randomised for each participants, as was the order of the three questions within each task.

The tasks labels in the table below were not seen by participants and refer to the type of alternative negation using in the tasks. Tasks UN1 to UN8 included a question using “un-” negation. Tasks IN1 to IN8 included a question using “iN-” negation. Tasks NO1 to NO8 included a question using “non-” negation. Tasks IM1 to IM8 included a question using implicit negation (based on an opposite term).

Approximately half the participants saw each question using “not” negation in the form ‘Out of the 100 people, how many would you not expect to be ###’ (as shown below). The rest of the participants saw these questions in the format ‘Out of the 100 people, how many would you expect to be not ###’.

*Table A1.1 – Tasks in chapter 2, experiment 1*

Task	Scenario	Questions
UN1	100 people joined a gym.  When inducting people into the gym, the instructors expect people to have different levels of fitness.	<ul style="list-style-type: none"> <li>• Out of the 100 people, how many would you expect to be fit?</li> <li>• Out of the 100 people, how many would you not expect to be fit?</li> <li>• Out of the 100 people, how many would you expect to be unfit?</li> </ul>
UN2	100 people applied for a job.  The person sifting the applications expects applicants to have different levels of qualifications relating to the job.	<ul style="list-style-type: none"> <li>• Out of the 100 people, how many would you expect to be qualified?</li> <li>• Out of the 100 people, how many would you not expect to be qualified?</li> <li>• Out of the 100 people, how many would you expect to be unqualified?</li> </ul>
UN3	100 people put their photos on a dating website.  A person browsing the site expects people to have different levels of attractiveness.	<ul style="list-style-type: none"> <li>• Out of the 100 people, how many would you expect to be attractive?</li> <li>• Out of the 100 people, how many would you not expect to be attractive?</li> <li>• Out of the 100 people, how many would you expect to be unattractive?</li> </ul>

Task	Scenario	Questions
UN4	<p>100 people lie on a new type of mattress in a shop.</p> <p>The sales assistant knows that different people will experience different levels of comfort on the mattress.</p>	<ul style="list-style-type: none"> <li>• Out of the 100 people, how many would you expect to be comfortable?</li> <li>• Out of the 100 people, how many would you not expect to be comfortable?</li> <li>• Out of the 100 people, how many would you expect to be uncomfortable?</li> </ul>
UN5	<p>100 people are selected to become part of juries for a series of trials.</p> <p>The judge knows that members of a jury will have different levels of biases based on their expectations.</p>	<ul style="list-style-type: none"> <li>• Out of the 100 people, how many would you expect to be biased?</li> <li>• Out of the 100 people, how many would you not expect to be not biased?</li> <li>• Out of the 100 people, how many would you expect to be unbiased?</li> </ul>
UN6	<p>100 people start on a series of creative writing courses.</p> <p>The tutors expect people on the course to demonstrate different levels of imagination as the course progresses.</p>	<ul style="list-style-type: none"> <li>• Out of the 100 people, how many would you expect to be imaginative?</li> <li>• Out of the 100 people, how many would you not expect to be imaginative?</li> <li>• Out of the 100 people, how many would you expect to be unimaginative?</li> </ul>
UN7	<p>100 people volunteer to take part in a medical trial.</p> <p>Doctors running the trial know that the volunteers will have different levels of suitability.</p>	<ul style="list-style-type: none"> <li>• Out of the 100 people, how many would you expect to be suitable?</li> <li>• Out of the 100 people, how many would you not expect to be suitable?</li> <li>• Out of the 100 people, how many would you expect to be unsuitable?</li> </ul>
UN8	<p>100 people are asked if they can attend a party.</p> <p>The organisers know that the people they invite will have different levels of availability.</p>	<ul style="list-style-type: none"> <li>• Out of the 100 people, how many would you expect to be available?</li> <li>• Out of the 100 people, how many would you not expect to be available?</li> <li>• Out of the 100 people, how many would you expect to be unavailable?</li> </ul>
IN1	<p>100 people are asked to change their shift patterns at work.</p> <p>The boss knows that their workers will have different levels of flexibility in being able to change shifts.</p>	<ul style="list-style-type: none"> <li>• Out of the 100 people, how many would you expect to be flexible?</li> <li>• Out of the 100 people, how many would you not expect to be flexible?</li> <li>• Out of the 100 people, how many would you expect to be inflexible?</li> </ul>

Task	Scenario	Questions
IN2	<p>100 people are asked for their views on an emotional issue.</p> <p>The questioners know that different people have different levels of rationality when considering emotional issues.</p>	<ul style="list-style-type: none"> <li>• Out of the 100 people, how many would you expect to be rational?</li> <li>• Out of the 100 people, how many would you not expect to be rational?</li> <li>• Out of the 100 people, how many would you expect to be irrational?</li> </ul>
IN3	<p>100 people have to give bad news to their staff.</p> <p>The staff know that they can expect different levels of directness depending on who their boss is.</p>	<ul style="list-style-type: none"> <li>• Out of the 100 people, how many would you expect to be direct?</li> <li>• Out of the 100 people, how many would you not expect to be direct?</li> <li>• Out of the 100 people, how many would you expect to be indirect?</li> </ul>
IN4	<p>100 people apply for a Government benefit.</p> <p>An assessor knows that applicants' levels of eligibility for the benefits will vary.</p>	<ul style="list-style-type: none"> <li>• Out of the 100 people, how many would you expect to be eligible?</li> <li>• Out of the 100 people, how many would you not expect to be eligible?</li> <li>• Out of the 100 people, how many would you expect to be ineligible?</li> </ul>
IN5	<p>100 people like to drink in a bar.</p> <p>The bartender knows that different customers will come to the bar with different levels of regularity.</p>	<ul style="list-style-type: none"> <li>• Out of the 100 people, how many would you expect to be regular?</li> <li>• Out of the 100 people, how many would you not expect to be regular?</li> <li>• Out of the 100 people, how many would you expect to be irregular?</li> </ul>
IN6	<p>100 people are voting in a national election.</p> <p>Returning officers understand that voters have different levels of significance in deciding the result depending on how and where they vote.</p>	<ul style="list-style-type: none"> <li>• Out of the 100 people, how many would you expect to be significant?</li> <li>• Out of the 100 people, how many would you not expect to be significant?</li> <li>• Out of the 100 people, how many would you expect to be insignificant?</li> </ul>

Task	Scenario	Questions
IN7	<p>100 people each make a post on an internet forum.</p> <p>The forum moderators know that different people will make posts with different levels of relevance to the subject at hand.</p>	<ul style="list-style-type: none"> <li>• Out of the 100 people, how many would you expect to be relevant?</li> <li>• Out of the 100 people, how many would you not expect to be relevant?</li> <li>• Out of the 100 people, how many would you expect to be irrelevant?</li> </ul>
IN8	<p>100 people are being considered by managers who plan a couple of redundancies.</p> <p>The managers know that their employees will feel different levels of job security under these conditions.</p>	<ul style="list-style-type: none"> <li>• Out of the 100 people, how many would you expect to be secure?</li> <li>• Out of the 100 people, how many would you not expect to be secure?</li> <li>• Out of the 100 people, how many would you expect to be insecure?</li> </ul>
NO1	<p>100 people attend a political debate.</p> <p>The debate chairperson knows that different members of the audience will have different levels of political partisanship.</p>	<ul style="list-style-type: none"> <li>• Out of the 100 people, how many would you expect to be partisan?</li> <li>• Out of the 100 people, how many would you not expect to be partisan?</li> <li>• Out of the 100 people, how many would you expect to be nonpartisan?</li> </ul>
NO2	<p>100 people watch a music video on television.</p> <p>The music channel schedulers know that viewers will have different levels of admiration for the star of the video.</p>	<ul style="list-style-type: none"> <li>• Out of the 100 people, how many would you expect to be admirers?</li> <li>• Out of the 100 people, how many would you not expect to be admirers?</li> <li>• Out of the 100 people, how many would you expect to be nonadmirers?</li> </ul>
NO3	<p>100 people try some DIY over a bank holiday weekend.</p> <p>A hardware store manager knows that people trying DIY have different levels of practicality.</p>	<ul style="list-style-type: none"> <li>• Out of the 100 people, how many would you expect to be practical?</li> <li>• Out of the 100 people, how many would you not expect to be practical?</li> <li>• Out of the 100 people, how many would you expect to be non practical?</li> </ul>

Task	Scenario	Questions
NO4	<p>100 people start a part time course lasting four years.</p> <p>Those running the course need to estimate how many people will drop out and know that people will have different levels of commitment to it.</p>	<ul style="list-style-type: none"> <li>• Out of the 100 people, how many would you expect to be committed?</li> <li>• Out of the 100 people, how many would you not expect to be committed?</li> <li>• Out of the 100 people, how many would you expect to be noncommitted?</li> </ul>
NO5	<p>100 people attend a protest.</p> <p>The police know that protesters will have different levels of aggressiveness.</p>	<ul style="list-style-type: none"> <li>• Out of the 100 people, how many would you expect to be aggressive?</li> <li>• Out of the 100 people, how many would you not expect to be aggressive?</li> <li>• Out of the 100 people, how many would you expect to be nonaggressive?</li> </ul>
NO6	<p>100 people are in a park in the summer.</p> <p>A doctor knows that different people will have different levels of allergy for the plants and flowers in the park.</p>	<ul style="list-style-type: none"> <li>• Out of the 100 people, how many would you expect to be allergic?</li> <li>• Out of the 100 people, how many would you not expect to be allergic?</li> <li>• Out of the 100 people, how many would you expect to be nonallergic?</li> </ul>
NO7	<p>100 people are planning weddings.</p> <p>A stallholder at a wedding show knows that people have like different levels of tradition at their wedding.</p>	<ul style="list-style-type: none"> <li>• Out of the 100 people, how many would you expect to be traditional?</li> <li>• Out of the 100 people, how many would you not expect to be traditional?</li> <li>• Out of the 100 people, how many would you expect to be nontraditional?</li> </ul>
NO8	<p>100 people take part in a casual tournament for a game.</p> <p>The judges know that people have different levels of expertise at the game and the tournament is open to beginners and veterans alike.</p>	<ul style="list-style-type: none"> <li>• Out of the 100 people, how many would you expect to be expert?</li> <li>• Out of the 100 people, how many would you not expect to be expert?</li> <li>• Out of the 100 people, how many would you expect to be nonexpert?</li> </ul>
IM1	<p>100 people are spending time in a park.</p> <p>A park manager knows that different people will have different levels of movement when they are in the park.</p>	<ul style="list-style-type: none"> <li>• Out of the 100 people, how many would you expect to be moving?</li> <li>• Out of the 100 people, how many would you not expect to be moving?</li> <li>• Out of the 100 people, how many would you expect to be still?</li> </ul>

Task	Scenario	Questions
IM2	<p>100 people are trying to climb a mountain with a small forest half way up.</p> <p>An expert knows that at a given point, different climbers will be at different levels relative to the forest.</p>	<ul style="list-style-type: none"> <li>• Out of the 100 people, how many would you expect to be above?</li> <li>• Out of the 100 people, how many would you not expect to be above?</li> <li>• Out of the 100 people, how many would you expect to be below?</li> </ul>
IM3	<p>100 people are going to visit the Doctor.</p> <p>The Doctor knows that different people will have different weights.</p>	<ul style="list-style-type: none"> <li>• Out of the 100 people, how many would you expect to be fat?</li> <li>• Out of the 100 people, how many would you not expect to be fat?</li> <li>• Out of the 100 people, how many would you expect to be slender?</li> </ul>
IM4	<p>100 people are asked about something they like to be secret from most people by close friends.</p> <p>Friends know that they can expect different levels of openness from different people.</p>	<ul style="list-style-type: none"> <li>• Out of the 100 people, how many would you expect to be open?</li> <li>• Out of the 100 people, how many would you not expect to be open?</li> <li>• Out of the 100 people, how many would you expect to be closed?</li> </ul>
IM5	<p>100 people are going to travel on an aeroplane.</p> <p>The designer of the plane's seats will know that different people traveling by air will be different sizes.</p>	<ul style="list-style-type: none"> <li>• Out of the 100 people, how many would you expect to be large?</li> <li>• Out of the 100 people, how many would you not expect to be large?</li> <li>• Out of the 100 people, how many would you expect to be small?</li> </ul>
IM6	<p>100 people are listed in order of their scores on an exam pass-list.</p> <p>The examiner knows that different people will be at different levels on the list.</p>	<ul style="list-style-type: none"> <li>• Out of the 100 people, how many would you expect to be top?</li> <li>• Out of the 100 people, how many would you not expect to be top?</li> <li>• Out of the 100 people, how many would you expect to be bottom?</li> </ul>
IM7	<p>100 people are running in a race.</p> <p>The race organiser knows that at a given point, different people will at different positions in the race.</p>	<ul style="list-style-type: none"> <li>• Out of the 100 people, how many would you expect to be in front?</li> <li>• Out of the 100 people, how many would you not expect to be in front?</li> <li>• Out of the 100 people, how many would you expect to be behind?</li> </ul>



Task	Scenario	Questions
IM8	100 people are down in a cave.  An experienced caver knows that different people will spend time at different depths in the cave.	<ul style="list-style-type: none"> <li>• Out of the 100 people, how many would you expect to be deep?</li> <li>• Out of the 100 people, how many would you not expect to be deep?</li> <li>• Out of the 100 people, how many would you expect to be shallow?</li> </ul>

## A1.2. Experiment 2

The second experiment in chapter 2 looked at whether negation using affixes “non-”, “un-” and “iN-” led to different probability ratings than negation using “not”. This experiment is reported in §2.4. The method is set out in §2.4.3 and the development of the materials is discussed in §2.4.3.3.

### A1.2.1. Task Instructions

The following text was provided to participants to explain the task:

*On each of the following pages, you will be given a scenario and asked to answer a question about what you might expect given the scenario. Your response to each question will be an answer between 0 and 100.*

*This is asking for your view, there are no right or wrong answers and no questions are trying to catch you out.*

*You should try to complete each question as quickly as possible – go with your first reaction.*

*When you are ready to begin, click on 'Next' for the first question.*

### A1.2.2. Task Questions

The following table sets out the five scenarios presented in this experiment. Each participant saw each of the scenarios in random order with one of the questions attached to each scenario (they saw one question in each of the five polarity conditions across the five scenarios). The tasks were presented in an online form with a scenario and the assigned question for that scenario on a single page.

The scenario labels are provided for convenient reference and were not provided to participants.

*Table A1.2 – Tasks in chapter 2, experiment 2*

Label	Scenario	Questions
A	<p>100 people live in their parents' house.</p> <p>- Some of those people pay a commercial rate of rent.</p> <p>- Some of those people also use a parent's car.</p> <p>- Some of those people have parents who are away for most of the year.</p> <p>- Some of those people also go on holiday with their parents.</p>	<ul style="list-style-type: none"> <li>• Out of the 100 people, how many would you expect to be dependent?</li> <li>• Out of the 100 people, how many would you not expect to be dependent?</li> <li>• Out of the 100 people, how many would you expect to be nondependent?</li> <li>• Out of the 100 people, how many would you expect to be independent?</li> <li>• Out of the 100 people, how many would you expect to be independent?</li> </ul>
B	<p>100 people work for local government.</p> <p>- Some of those people like to read in their spare time.</p> <p>- Some of those people will often be out of the office, meeting contractors.</p> <p>- Some of those people will spend most of their working time using the computer and phone.</p> <p>- Some of those people will be regarded as fit by their co-workers.</p>	<ul style="list-style-type: none"> <li>• Out of the 100 people, how many would you expect to be active?</li> <li>• Out of the 100 people, how many would you not expect to be active?</li> <li>• Out of the 100 people, how many would you expect to be nonactive?</li> <li>• Out of the 100 people, how many would you expect to be inactive?</li> <li>• Out of the 100 people, how many would you expect to be inactive?</li> </ul>

Label	Scenario	Questions
C	<p>100 people get up early on a Sunday.</p> <ul style="list-style-type: none"> <li>- Some of those people will be scientists.</li> <li>- Some of those people will have been taken to church every week when they were young.</li> <li>- Some of those people will not be superstitious.</li> <li>- Some of those people will be filled with awe when visiting large cathedrals.</li> </ul>	<ul style="list-style-type: none"> <li>• Out of the 100 people, how many would you expect to be religious?</li> <li>• Out of the 100 people, how many would you not expect to be religious?</li> <li>• Out of the 100 people, how many would you expect to be nonreligious?</li> <li>• Out of the 100 people, how many would you expect to be unreligious?</li> <li>• Out of the 100 people, how many would you expect to be irreligious?</li> </ul>
D	<p>100 people had to complain about appalling service in a shop.</p> <ul style="list-style-type: none"> <li>- Some of those people were in a hurry.</li> <li>- Some of those people don't like to be aggressive.</li> <li>- Some of those people had been made very angry by the service.</li> <li>- Some of those people will try to stay calm.</li> </ul>	<ul style="list-style-type: none"> <li>• Out of the 100 people, how many would you expect to be polite?</li> <li>• Out of the 100 people, how many would you not expect to be polite?</li> <li>• Out of the 100 people, how many would you expect to be nonpolite?</li> <li>• Out of the 100 people, how many would you expect to be unpolite?</li> <li>• Out of the 100 people, how many would you expect to be impolite?</li> </ul>
E	<p>100 people speak to a police officer.</p> <ul style="list-style-type: none"> <li>- Some of those people will have been violent.</li> <li>- Some of those people will feel deferential.</li> <li>- Some of those people will remain calm.</li> <li>- Some of those people will dislike the police.</li> </ul>	<ul style="list-style-type: none"> <li>• Out of the 100 people, how many would you expect to be offensive?</li> <li>• Out of the 100 people, how many would you not expect to be offensive?</li> <li>• Out of the 100 people, how many would you expect to be nonoffensive?</li> <li>• Out of the 100 people, how many would you expect to be unoffensive?</li> <li>• Out of the 100 people, how many would you expect to be inoffensive?</li> </ul>

### A1.3. Experiment 3

The third and final experiment in chapter 3 first looked at whether negation using affix “un-” led to different probability ratings than negation using “not”. A second part

sought plausibility ratings for conditionals using antecedent and consequent terms which were systematically varied between affirmative, “not” negation and “un-” negated conditions. This experiment is reported in §2.5. The method is set out in §2.5.2 and the development of the materials is discussed in §2.5.2.3.

### **A1.3.1. Task Instructions for Part 1**

The following text was provided to participants to explain the first set of tasks:

*In the first part of this survey, you will be given a scenario and asked to answer a question about what you might expect given the scenario. Your response to each question will be an answer between 0 and 100.*

*This is asking for your view, there are no right or wrong answers and no questions are trying to catch you out.*

*You should try to complete each question as quickly as possible – go with your first reaction.*

*When you are ready to begin, click on 'Next' for the first question.*

### **A1.3.2. Task Questions for Part 1**

The following table sets out the nine scenarios presented in part 1 of this experiment. Each participant saw each of the scenarios in random order with one of the questions attached to the scenario (randomly determined). The tasks were presented in an online form with a scenario and its question on a single page.

The scenario labels are provided for convenient reference and were not provided to participants.

*Table A1.3 – Tasks in chapter 2, experiment 3, part 1*

Label	Scenario	Questions
1	<p>100 people are asked if they believe in a newspapers investigative report.</p> <p>- Many of those people will be convinced by the evidence.</p> <p>- A few of those people never trust reports.</p> <p>- Several of those people normally accept such reports without much consideration.</p> <p>- Some of those people will want to challenge the evidence.</p>	<ul style="list-style-type: none"> <li>• Out of the 100 people, how many would you expect to be believing?</li> <li>• Out of the 100 people, how many would you expect to be not believing?</li> <li>• Out of the 100 people, how many would you expect to be unbelieving?</li> </ul>
2	<p>100 people were asked how they felt about their neighbours.</p> <p>- Many of those people often spoke with their neighbours.</p> <p>- A few of those people actively disliked their neighbours.</p> <p>- Several of those people sometimes socialised with their neighbours.</p> <p>- Some of those people didn't often see their neighbours.</p>	<ul style="list-style-type: none"> <li>• Out of the 100 people, how many would you expect to say their neighbours were friendly?</li> <li>• Out of the 100 people, how many would you expect to say their neighbours were not friendly?</li> <li>• Out of the 100 people, how many would you expect to say their neighbours were unfriendly?</li> </ul>
3	<p>100 people where asked how they felt about a promotional gift when purchasing another item.</p> <p>- Many of those people liked the promotional gift.</p> <p>- A few of those people were annoyed by the promotional gift.</p> <p>- Several of those people had wanted the promotional gift.</p> <p>- Some of those people would have preferred not to have received the promotional gift.</p>	<ul style="list-style-type: none"> <li>• Out of the 100 people, how many would you expect to be grateful?</li> <li>• Out of the 100 people, how many would you expect to be not grateful?</li> <li>• Out of the 100 people, how many would you expect to be ungrateful?</li> </ul>

Label	Scenario	Questions
4	<p>100 people are asked about their mood.</p> <ul style="list-style-type: none"> <li>- Many of those people had been doing something they enjoyed.</li> <li>- A few of those people were upset about something earlier that day.</li> <li>- Several of those people had recently received good news.</li> <li>- Some of those people found what they were doing a bit tedious.</li> </ul>	<ul style="list-style-type: none"> <li>• Out of the 100 people, how many would you expect to be happy?</li> <li>• Out of the 100 people, how many would you expect to be not happy?</li> <li>• Out of the 100 people, how many would you expect to be unhappy?</li> </ul>
5	<p>100 people are asked what they plan for Valentines Day.</p> <ul style="list-style-type: none"> <li>- Many of those people want to do something nice with their partners.</li> <li>- A few of those people will upset their partners on Valentines Day.</li> <li>- Several of those people spend a lot of time planning their evening.</li> <li>- Some of those people are cynical about Valentines Day.</li> </ul>	<ul style="list-style-type: none"> <li>• Out of the 100 people, how many would you expect to be romantic?</li> <li>• Out of the 100 people, how many would you expect to be not romantic?</li> <li>• Out of the 100 people, how many would you expect to be unromantic?</li> </ul>
6	<p>100 people have their dress-sense assessed.</p> <ul style="list-style-type: none"> <li>- Many of those people will take care in choosing their clothes.</li> <li>- A few of those people will dress in a way that others consider appalling.</li> <li>- Several of those people will have taste others admire.</li> <li>- Some of those people will be thought to look okay but not particularly good or bad.</li> </ul>	<ul style="list-style-type: none"> <li>• Out of the 100 people, how many would you expect to be stylish?</li> <li>• Out of the 100 people, how many would you expect to be not stylish?</li> <li>• Out of the 100 people, how many would you expect to be unstylish?</li> </ul>

Label	Scenario	Questions
7	<p>100 people are asked how they would make a difficult decision affecting other people.</p> <ul style="list-style-type: none"> <li>- Many of those people would make a sound decision.</li> <li>- A few of those people would make a reckless decision.</li> <li>- Several of those people would base their decision on extensive experience.</li> <li>- Some of those people would only have limited knowledge on which to base their decision.</li> </ul>	<ul style="list-style-type: none"> <li>• Out of the 100 people, how many would you expect to be wise?</li> <li>• Out of the 100 people, how many would you expect to be not wise?</li> <li>• Out of the 100 people, how many would you expect to be unwise?</li> </ul>
8	<p>100 people are asked how they would consider the guilt of a suspect given conflicting evidence.</p> <ul style="list-style-type: none"> <li>- Many of those people are willing to give the benefit of the doubt to the suspect.</li> <li>- A few of those people assume the suspect was guilty and look at how the evidence in their favour could be mitigated.</li> <li>- Several of those people would consider the suspect innocent until proven guilty.</li> <li>- Some of those people admit their reaction would depend on the nature of the crime and a critical view of all the evidence.</li> </ul>	<ul style="list-style-type: none"> <li>• Out of the 100 people, how many would you expect to be charitable?</li> <li>• Out of the 100 people, how many would you expect to be not charitable?</li> <li>• Out of the 100 people, how many would you expect to be uncharitable?</li> </ul>

Label	Scenario	Questions
9	<p>100 people are asked how they behave in a room full of new people.</p> <ul style="list-style-type: none"> <li>- Many of those people would try to make conversation.</li> <li>- A few of those people would find the situation so difficult they would want to leave the room.</li> <li>- Several of those people would want to use the opportunity to make new friends.</li> <li>- Some of those people would stay quiet.</li> <li>- 00 people had to complain about appalling service in a shop.</li> <li>- Some of those people were in a hurry.</li> <li>- Some of those people don't like to be aggressive.</li> <li>- Some of those people had been made very angry by the service.</li> <li>- Some of those people will try to stay calm.</li> </ul>	<ul style="list-style-type: none"> <li>• Out of the 100 people, how many would you expect to be social?</li> <li>• Out of the 100 people, how many would you expect to be not social?</li> <li>• Out of the 100 people, how many would you expect to be unsocial?</li> </ul>

### **A1.3.3. Task Instructions for Part 2**

The following text was provided to participants to explain the second part of this experiment:

*In each question in the final part of this survey, you will be given a scenario and a rule. You will be asked indicate how plausible you think the rule is given the scenario. You should select your response from the available options.*

*This is asking for your view, there are no right or wrong answers and no questions are trying to catch you out.*

*You should try to complete each question as quickly as possible – go with your first reaction.*

*When you are ready to begin, click on 'Next' for the first question.*



#### **A1.3.4. Task Questions for Part 2**

The following table sets out the four scenarios presented in part 2 of this experiment. Each participant saw each of the scenarios in random order. Each scenario was presented with one of the nine related conditionals (randomly determined for each participant) on a single page of an online form with the experimental question. Each scenario page had the following format:

*[scenario]*

*An observer looks at the 100 people and suggests the following rule:*

*[conditional]*

*Given the information provided, how plausible do you think this rule is?*

*[Options provided: Completely implausible/Very implausible/Implausible/A little implausible/A little Plausible/Plausible/Very plausible/Completely plausible]*

The scenario labels are provided for convenient reference and were not provided to participants.

*Table A1.4 – Tasks in chapter 2, experiment 3, part 2*

Label	Scenario	Conditional
1	<p>100 people were asked how they felt about their neighbours.</p> <p>- Many of those people often spoke with their neighbours.</p> <p>- A few of those people actively disliked their neighbours.</p> <p>- Several of those people sometimes socialised with their neighbours.</p> <p>- Some of those people didn't often see their neighbours.</p> <p>The same 100 people are asked how they behave in a room full of new people.</p> <p>- Many of those people would try to make conversation.</p> <p>- A few of those people would find the situation so difficult they would want to leave the room.</p> <p>- Several of those people would want to use the opportunity to make new friends.</p> <p>- Some of those people would stay quiet.</p>	<ul style="list-style-type: none"> <li>• If a person is friendly, then the person is social</li> <li>• If a person is friendly, then the person is not social</li> <li>• If a person is friendly, then the person is unsocial</li> <li>• If a person is not friendly, then the person is social</li> <li>• If a person is not friendly, then the person is not social</li> <li>• If a person is not friendly, then the person is unsocial</li> <li>• If a person is unfriendly, then the person is social</li> <li>• If a person is unfriendly, then the person is not social</li> <li>• If a person is unfriendly, then the person is unsocial</li> </ul>

Label	Scenario	Conditional
2	<p>100 people were asked how they felt about their neighbours.</p> <p>- Many of those people often spoke with their neighbours.</p> <p>- A few of those people actively disliked their neighbours.</p> <p>- Several of those people sometimes socialised with their neighbours.</p> <p>- Some of those people didn't often see their neighbours.</p> <p>The same 100 people are asked how they behave in a room full of new people.</p> <p>- Many of those people would try to make conversation.</p> <p>- A few of those people would find the situation so difficult they would want to leave the room.</p> <p>- Several of those people would want to use the opportunity to make new friends.</p> <p>- Some of those people would stay quiet.</p>	<ul style="list-style-type: none"> <li>• If a person is friendly, then the person is social</li> <li>• If a person is friendly, then the person is not social</li> <li>• If a person is friendly, then the person is unsocial</li> <li>• If a person is not friendly, then the person is social</li> <li>• If a person is not friendly, then the person is not social</li> <li>• If a person is not friendly, then the person is unsocial</li> <li>• If a person is unfriendly, then the person is social</li> <li>• If a person is unfriendly, then the person is not social</li> <li>• If a person is unfriendly, then the person is unsocial</li> </ul>

Label	Scenario	Conditional
3	<p>100 people where asked how they felt about a promotional gift when purchasing another item.</p> <p>- Many of those people liked the promotional gift.</p> <p>- A few of those people were annoyed by the promotional gift.</p> <p>- Several of those people had wanted the promotional gift.</p> <p>- Some of those people would have preferred not to have received the promotional gift.</p> <p>The same 100 people are asked about their mood.</p> <p>- Many of those people had been doing something they enjoyed.</p> <p>- A few of those people were upset about something earlier that day.</p> <p>- Several of those people had recently received good news.</p> <p>- Some of those people found what they were doing a bit tedious.</p>	<ul style="list-style-type: none"> <li>• If a person is grateful, then the person is happy</li> <li>• If a person is grateful, then the person is not happy</li> <li>• If a person is grateful, then the person is unhappy</li> <li>• If a person is not grateful, then the person is happy</li> <li>• If a person is not grateful, then the person is not happy</li> <li>• If a person is not grateful, then the person is unhappy</li> <li>• If a person is ungrateful, then the person is happy</li> <li>• If a person is ungrateful, then the person is not happy</li> <li>• If a person is ungrateful, then the person is unhappy</li> </ul>

Label	Scenario	Conditional
4	<p>100 people are asked what they plan for Valentines Day.</p> <p>- Many of those people want to do something nice with their partners.</p> <p>- A few of those people will upset their partners on Valentines Day.</p> <p>- Several of those people spend a lot of time planning their evening.</p> <p>- Some of those people are cynical about Valentines Day.</p> <p>The same 100 people have their dress-sense assessed.</p> <p>- Many of those people will take care in choosing their clothes.</p> <p>- A few of those people will dress in a way that others consider appalling.</p> <p>- Several of those people will have taste others admire.</p> <p>- Some of those people will be thought to look okay but not particularly good or bad.</p>	<ul style="list-style-type: none"> <li>• If a person is romantic, then the person is stylish</li> <li>• If a person is romantic, then the person is not stylish</li> <li>• If a person is romantic, then the person is unstylish</li> <li>• If a person is not romantic, then the person is stylish</li> <li>• If a person is not romantic, then the person is not stylish</li> <li>• If a person is not romantic, then the person is unstylish</li> <li>• If a person is unromantic, then the person is stylish</li> <li>• If a person is unromantic, then the person is not stylish</li> <li>• If a person is unromantic, then the person is unstylish</li> </ul>

## **Appendix 2. Materials Used in Chapter 3 Experiments**

This appendix contains the experimental materials used in each of the four experiments reported in chapter 3. These used an extension of Evans' (1977) conditional inference paradigm to compare two types of negation (using "not" and affix "un-"). The experiments looked at whether responses reflected the different probabilistic implications of those negative forms.

### **A2.1. Experiment 1**

The first experiment in chapter 3 first looked at how people endorsed inferences using conditionals when the antecedent term was affirmative, used "not" negation and used "un-" negation. A second part sought plausibility ratings for conditionals using antecedent and consequent terms which were systematically varied between affirmative, "not" negation and "un-" negated conditions. This experiment is reported in §3.3. The method is set out in §3.3.2 and the development of the materials is discussed in §3.3.2.3.

#### **A2.1.1. Task Instructions for Part 1**

The following text was provided to participants to explain the first set of tasks:

*The first part of this study consists of a series of pages which each have a conditional inference task.*

*For each task you will be given two premises (a relationship such as "If Adrian has a hangover tomorrow then he was drinking" and statement like "Adrian has a hangover tomorrow") and a conclusion (like "Therefore Adrian was drinking").*

*You should press "A" if you believe the conclusion follows from the two premises and "L" if you do not believe the conclusion follows from the two premises.*

*You should try to be accurate as you respond to each task as quickly as possible.*

*Press any key when you are ready to start.*

### A2.1.2. Task Questions for Part 1

Participants were presented with 144 inferences, one to a screen, in random order. The 144 inferences were made up of twelve scenarios each presented in four different inference forms tested (MP, DA, AC, MT) each using three different antecedent polarity conditions (affirmative, “not” negated, “un-” negated). The basic forms presented for each task of each inference type were as follows:

- MP
  - *If [s1] is [n1][s2] [s3] then [s4] will be [s5].*
  - *[s1] is [n1][s2] [s3].*
  - *Therefore [s1] will be [s5].*
- DA
  - *If [s1] is [n1][s2] [s3] then [s4] will be [s5].*
  - *[s1] is [n2] [s2] [s3].*
  - *Therefore [s1] won't be [s5].*
- AC
  - *If [s1] is [n1][s2] [s3] then [s4] will be [s5].*
  - *[s1] will be [s5].*
  - *Therefore [s1] is [n1][s2] [s3].*
- MT
  - *If [s1] is [n1][s2] [s3] then [s4] will be [s5].*
  - *[s1] won't be [s5].*
  - *Therefore [s1] is [n2] [s2] [s3].*

The placeholders [s1] to [s5] were varied to create the twelve different scenario conditions. The following table sets out the content used in these placeholders for each of the twelve scenarios.

*Table A2.1 – Scenarios in chapter 3, experiment 1, part 1*

Scenario	[s1]	[s2]	[s3]	[s4]	[s5]
1	Andrew	predictable	in the kitchen tomorrow	he	clean
2	Melanie	stable	at home on Sunday	she	tidy
3	Peter	enthusiastic	about visiting his brother next month	he	helpful
4	Susan	happy	watching the programme on Wednesday	she	interested
5	Paul	believable	at the trial next week	he	just
6	Brenda	informed	about the issue being discussed today	she	concerned
7	Stephen	responsive	to Karen on her birthday	he	settled

Scenario	[s1]	[s2]	[s3]	[s4]	[s5]
8	Joanna	kind	to her friends at the party this evening	she	forgiving
9	Alan	popular	after school on Tuesday	he	friendly
10	Sarah	reliable	on Friday morning at work	she	sympathetic
11	Mark	successful	in the gym on Saturday	he	critical
12	Liz	conventional	in planning her holiday for this summer	she	enthusiastic

The placeholders [n1] and [n2] were varied to create the three antecedent polarity conditions. The following table sets out the content used in these placeholders for each of the three conditions.

*Table A2.2 – Polarity conditions in chapter 3, experiment 1, part 1*

Polarity	[n1]	[n2]
Affirmative		n't
"not" negated	not	
"un-" negated	un-	

### **A2.1.3. Task Instructions for Part 2**

The following text was provided to participants to explain the second set of tasks:

*The final part of this study consists of a series of pages which ask you to assess the plausibility of a relationship.*

*On each page you will be given a relationship (for example "If there is a fly in the soup then the waiter takes it away"). There is also a line which represents 'entirely plausible' on the left and 'entirely implausible' on the right.*

*You should click on the line at the point that best represents how plausible you think the relationship is.*

*You should try and respond to each task as quickly as possible.*

*Please click the mouse when you are ready to start.*



#### **A2.1.4. Task Questions for Part 2**

Participants were presented with 108 conditionals, one to a screen, in random order. The 108 inferences were made up of twelve scenarios each presented in nine polarity conditions. The nine polarity conditions were made up by systematically varying both antecedent and consequent polarity across three conditions (affirmative, “not” negated, “un-” negated). The basic conditional form presented for each task was as follows:

*If [s1] is [n1][s2] [s3] then [s4] will [n2][s5]*

The placeholders [s1] to [s5] were varied to create the twelve different scenario conditions. The scenarios were the same as for part 1 and so the content was as set out in table A2.1 (above).

The placeholders [n1] and [n2] were varied to create the nine antecedent polarity conditions. The following table sets out the content used in these placeholders for each of the three conditions.

*Table A2.3 – Polarity conditions in chapter 3, experiment 1, part 2*

Antecedent Polarity	Consequent Polarity	[n1]	[n2]
Affirmative	Affirmative		be_
Affirmative	“not” negated		not be_
Affirmative	“un-” negated		be un-
“not” negated	Affirmative	not_	be_
“not” negated	“not” negated	not_	not be_
“not” negated	“un-” negated	not_	be un-
“un-” negated	Affirmative	un-	be_
“un-” negated	“not” negated	un-	not be_
“un-” negated	“un-” negated	un-	be un-

#### **A2.2. Experiment 2**

The second experiment in chapter 3 first looked at how people endorsed inferences using conditionals with the antecedent and consequent terms being systematically varied between affirmative “not” negated and used “un-” negated conditions. This

experiment is reported in §3.4. The method is set out in §3.4.2 and the development of the materials is discussed in §3.4.2.3.

### **A2.2.1. Task Instructions**

The following text was provided to participants to explain tasks:

*The study consists of a series of pages, with one task on each page.*

*For each task you will be given two premises (a relationship such as "If Adrian has a hangover tomorrow then he was drinking" and statement like "Adrian has a hangover tomorrow") and a conclusion (like "Therefore Adrian was drinking").*

*You should indicate whether or not you think the conclusion follows from the two premises.*

*You should try to answer each question as quickly as possible - go with your first reaction.*

*This study is looking at your perception. There are no right or wrong answers and no question is trying to catch you out.*

*Over the next pages there are three practice questions for you to complete before moving onto the study questions.*

*Click 'Next' when you are ready to start the practice questions.*

After the practice questions the following instructions were presented and before the experimental tasks, the instructions were repeated as follows:

*You have now completed the practice questions. Before you move onto the study questions, here is a reminder of the instructions.*

*The study consists of a series of pages, with one task on each page.*

*For each task you will be given two premises (a relationship such as "If Adrian has a hangover tomorrow then he was drinking" and statement like "Adrian has a hangover tomorrow") and a conclusion (like "Therefore Adrian was drinking").*

*You should indicate whether or not you think the conclusion follows from the two premises.*

*You should try to answer each question as quickly as possible - go with your first reaction.*

*This study is looking at your perception. There are no right or wrong answers and no question is trying to catch you out.*

*Click 'Next' when you are ready to start the study questions.*

### **A2.2.2.Practice Questions**

After the initial instructions, the following three practice inference tasks were presented, each on a single screen:

- Practice 1
  - *If Peter is not enthusiastic about his visit then he is helpful.*
  - *Peter is not enthusiastic about his visit.*
  - *Therefore Peter is helpful.*
- Practice 2
  - *If Susan is not happy on Wednesday then she is uninterested.*
  - *Susan is interested.*
  - *Therefore Susan is happy on Wednesday.*
- Practice 3
  - *If Paul is unbelievable in court then he is unjust.*
  - *Paul is believable in court.*
  - *Therefore Paul is just.*

Below each inference, participants were given the following two options (in random order). Participants were required to select one and click a ‘Next’ button to move on.

- *The conclusion follows from the premises.*
- *The conclusion does not follow from the premises.*

### **A2.2.3. Task Questions**

Participants were presented with 108 inferences, one to a screen, in random order. The 108 inferences were made up of three scenarios each presented in four difference inference forms tested (MP, DA, AC, MT) each presented in nine polarity conditions. The nine polarity conditions were made up by systematically varying both antecedent and consequent polarity across three conditions (affirmative, “not” negated, “un-” negated). The basic forms presented for each task of each inference type were as follows:

- MP
  - *If [s1] is [n1][s2] [s3] then [s4] will [n2][s5].*
  - *[s1] is [n1][s2] [s3].*
  - *Therefore [s1] will [n2][s5].*
- DA
  - *If [s1] is [n1][s2] [s3] then [s4] will [n2][s5].*
  - *[s1] is[n3] [s2] [s3].*
  - *Therefore [s1] [n4] be [s5].*
- AC
  - *If [s1] is [n1][s2] [s3] then [s4] will [n2][s5].*
  - *[s1] will [n2][s5].*
  - *Therefore [s1] is [n1][s2] [s3].*
- MT
  - *If [s1] is [n1][s2] [s3] then [s4] will [n2][s5].*
  - *[s1] [n4] be [s5].*
  - *Therefore [s1] is[n3] [s2] [s3].*

Like the practice questions, the following two options were presented below each inference (in random order) and participants were required to select one and click a ‘Next’ button to move on:

- *The conclusion follows from the premises.*
- *The conclusion does not follow from the premises.*

The placeholders [s1] to [s5] were varied to create the different scenario conditions. The following table sets out the content used in these placeholders for each of the scenarios.

*Table A2.4 – Scenarios in chapter 3, experiment 2*

Scenario	[s1]	[s2]	[s3]	[s4]	[s5]
Experimental 1	Liz	conventional	in planning her holiday for this summer	she	enthusiastic
Experimental 2	Brenda	informed	about the issue being discussed today	she	concerned

A further 72 non-experimental distraction tasks were generated using additional scenario information in the placeholders from experiment 1 (see §A2.1.2).

The placeholders [n1] to [n4] were varied to create the three antecedent polarity conditions. The following table sets out the content used in these placeholders for each of the three conditions.

*Table A2.5 – Polarity conditions in chapter 3, experiment 2*

Antecedent Polarity	Consequent Polarity	[n1]	[n2]	[n3]	[n4]
Affirmative	Affirmative		be_	n’t	won’t
Affirmative	“not” negated				not be_
Affirmative	“un-” negated		be un-	n’t	will
“not” negated	Affirmative	not_	be_		won’t
“not” negated	“not” negated	not_	not_	not be_	
“not” negated	“un-” negated	not_	be un-		will
“un-” negated	Affirmative	un-	be_		won’t
“un-” negated	“not” negated	un-	not be_		will
“un-” negated	“un-” negated	un-	be un-		will

### **A2.3. Experiments 3 and 4**

The final two experiments in chapter 3 looked at how people endorsed inferences using conditionals when the antecedent term was affirmative, used “not” negation and used “un-” negation. Information about participants perception of the plausibility of the conditionals and some propositions seen was also collected. The materials are presented together here because both experiments used the same set of inference tasks in different experimental designs.

Experiment 3 is reported in §3.5. The method is set out in §3.5.2 and the development of the materials is discussed in §3.5.2.3.

Experiment 4 is reported in §3.6. The method is set out in §3.6.2 and the development of the materials is discussed in §3.6.2.3.

#### **A2.3.1. Task Instructions for Experiments 3 and 4, Part 1**

The following text was provided to participants in experiment 3 to explain the inference tasks:

*For each of the four questions that follows, you will be given a scenario, a rule, a statement related to the scenario and a conclusion. You will be asked indicate whether you agree with the conclusion given the rule and statement. You should select your response from the available options.*

*After these four questions, four further questions will provide a scenario and a rule and ask you a question about how plausible you think the rule is given the scenario. You enter your response as a numerical answer between 0 and 100.*

*All of the questions are asking for your view, there are no right or wrong answers and no questions are trying to catch you out.*

*You should try to complete each question as quickly as possible – go with your first reaction.*

*When you are ready to begin, click on 'Next' for the first question.*

The following text was provided to participants in experiment 4 to explain the inference tasks:

*The first part of the survey will start on the next page.*

*On each page in this first part, you will be given a scenario, a rule, a statement related to the scenario and a conclusion. You will be asked indicate whether you agree with the conclusion given the rule and statement. You should select your response from the available options and click next to move onto the next question.*

*The scenarios given are each used in multiple questions. However, the rule, statement and conclusion change in each question so you should pay attention to this.*

*After this part of the survey is complete, the second part will consist of further questions which provide a scenario and a rule or statement and ask you a question about how plausible you think the rule or statement is given the scenario. You enter your response as a numerical answer between 0 and 100. All of the questions are asking for your view, there are no right or wrong answers and no questions are trying to catch you out.*

*You should try to complete each question as quickly as possible – go with your first reaction.*

*When you are ready to begin, click on 'Next' for the first question.*

### **A2.3.2. Task Questions for Experiments 3 and 4, Part 1**

144 inferences were prepared for experiments 3 and 4. The 144 inferences were made up of four scenarios each presented in four difference inference forms tested (MP, DA, AC, MT) each presented in nine polarity conditions. The nine polarity conditions were made up by systematically varying both antecedent and consequent polarity across three conditions (affirmative, “not” negated, “un-” negated).

In experiment 3, participants saw only four of the 144 inference tasks each. The four tasks they saw included one of each scenario and one of each inference type (antecedent and consequent polarity conditions were random for each inference seen). The four questions seen were presented in random order.

In experiment 4, participants saw all 144 inference tasks in random order.

Each task was presented one to a screen in the following form:

[scenario introduction]

*An observer looks at the 100 people and suggests the following rule:*

[conditional]

*Given the rule, if a person from the 100 people was found meeting the following condition:*

[second premise]

*Would you agree with the following conclusion?*

[conclusion]

The following eight options were presented below each inference (in order) and participants were required to select one and click a 'Next' button to move on:

- *Completely disagree*
- *Strongly disagree*
- *Disagree*
- *Disagree a little*
- *Agree a little* *Agree*
- *Strongly agree*
- *Completely agree*



The placeholder [scenario introduction] were varied for each of the four scenario conditions. The following table sets out the content used in this placeholder for each of the four scenarios.

*Table A2.6 – Scenario introductions in chapter 3, experiments 3 and 4*

Scenario	[scenario introduction]
1	<p>100 people are asked if they believe in a newspapers investigative report.</p> <ul style="list-style-type: none"> <li>- Many of those people will be convinced by the evidence.</li> <li>- A few of those people never trust reports.</li> <li>- Several of those people normally accept such reports without much consideration.</li> <li>- Some of those people will want to challenge the evidence.</li> </ul> <p>The same 100 people are also asked what they plan for Valentines Day.</p> <ul style="list-style-type: none"> <li>- Many of those people want to do something nice with their partners.</li> <li>- A few of those people will upset their partners on Valentines Day.</li> <li>- Several of those people spend a lot of time planning their evening.</li> <li>- Some of those people are cynical about Valentines Day.</li> </ul>
2	<p>100 people were asked how they felt about their neighbours.</p> <ul style="list-style-type: none"> <li>- Many of those people often spoke with their neighbours.</li> <li>- A few of those people actively disliked their neighbours.</li> <li>- Several of those people sometimes socialised with their neighbours.</li> <li>- Some of those people didn't often see their neighbours.</li> </ul> <p>The same 100 people are asked how they behave in a room full of new people.</p> <ul style="list-style-type: none"> <li>- Many of those people would try to make conversation.</li> <li>- A few of those people would find the situation so difficult they would want to leave the room.</li> <li>- Several of those people would want to use the opportunity to make new friends.</li> <li>- Some of those people would stay quiet.</li> </ul>

Scenario	[scenario introduction]
3	<p>100 people where asked how they felt about a promotional gift when purchasing another item.</p> <ul style="list-style-type: none"> <li>- Many of those people liked the promotional gift.</li> <li>- A few of those people were annoyed by the promotional gift.</li> <li>- Several of those people had wanted the promotional gift.</li> <li>- Some of those people would have preferred not to have received the promotional gift.</li> </ul> <p>The same 100 people are asked about their mood.</p> <ul style="list-style-type: none"> <li>- Many of those people had been doing something they enjoyed.</li> <li>- A few of those people were upset about something earlier that day.</li> <li>- Several of those people had recently received good news.</li> <li>- Some of those people found what they were doing a bit tedious.</li> </ul>
4	<p>100 people have their dress-sense assessed.</p> <ul style="list-style-type: none"> <li>- Many of those people will take care in choosing their clothes.</li> <li>- A few of those people will dress in a way that others consider appalling.</li> <li>- Several of those people will have taste others admire.</li> <li>- Some of those people will be thought to look okay but not particularly good or bad.</li> </ul> <p>The same 100 people are asked how they would make a difficult decision affecting other people.</p> <ul style="list-style-type: none"> <li>- Many of those people would make a sound decision.</li> <li>- A few of those people would make a reckless decision.</li> <li>- Several of those people would base their decision on extensive experience.</li> <li>- Some of those people would only have limited knowledge on which to base their decision.</li> </ul>

The placeholders [conditional], [second premise] and [conclusion] were varied between scenario, inference type, antecedent polarity and consequent polarity conditions. The following table sets out the content used in these placeholders for each condition.

*Table A2.7 – Inference tasks in chapter 3, experiments 3 and 4, part 1*

Scenario / Inference Type / Antecedent Polarity / Consequent Polarity	[conditional]	[second premise]	[conclusion]
1 / MP / Affirmative / Affirmative	If a person is believing, then the person is romantic	The person is believing	The person is romantic
1 / MP / Affirmative / “not” negated	If a person is believing, then the person is not romantic	The person is believing	The person is not romantic
1 / MP / Affirmative / “un-” negated	If a person is believing, then the person is unromantic	The person is believing	The person is unromantic
1 / MP / “not” negated / Affirmative	If a person is not believing, then the person is romantic	The person is not believing	The person is romantic
1 / MP / “not” negated / “not” negated	If a person is not believing, then the person is not romantic	The person is not believing	The person is not romantic
1 / MP / “not” negated / “un-” negated	If a person is not believing, then the person is unromantic	The person is not believing	The person is unromantic
1 / MP / “un-” negated / Affirmative	If a person is unbelieving, then the person is romantic	The person is unbelieving	The person is romantic
1 / MP / “un-” negated / “not” negated	If a person is unbelieving, then the person is not romantic	The person is unbelieving	The person is not romantic
1 / MP / “un-” negated / “un-” negated	If a person is unbelieving, then the person is unromantic	The person is unbelieving	The person is unromantic
1 / DA / Affirmative / Affirmative	If a person is believing, then the person is romantic	The person is not believing	The person is not romantic

Scenario / Inference Type / Antecedent Polarity / Consequent Polarity	[conditional]	[second premise]	[conclusion]
1 / DA / Affirmative / "not" negated	If a person is believing, then the person is not romantic	The person is not believing	The person is romantic
1 / DA / Affirmative / "un-" negated	If a person is believing, then the person is unromantic	The person is unbelieving	The person is romantic
1 / DA / "not" negated / Affirmative	If a person is not believing, then the person is romantic	The person is believing	The person is not romantic
1 / DA / "not" negated / "not" negated	If a person is not believing, then the person is not romantic	The person is believing	The person is romantic
1 / DA / "not" negated / "un-" negated	If a person is not believing, then the person is unromantic	The person is believing	The person is romantic
1 / DA / "un-" negated / Affirmative	If a person is unbelieving, then the person is romantic	The person is believing	The person is unromantic
1 / DA / "un-" negated / "not" negated	If a person is unbelieving, then the person is not romantic	The person is believing	The person is romantic
1 / DA / "un-" negated / "un-" negated	If a person is unbelieving, then the person is unromantic	The person is believing	The person is romantic
1 / AC / Affirmative / Affirmative	If a person is believing, then the person is romantic	The person is romantic	The person is believing
1 / AC / Affirmative / "not" negated	If a person is believing, then the person is not romantic	The person is not romantic	The person is believing

Scenario / Inference Type / Antecedent Polarity / Consequent Polarity	[conditional]	[second premise]	[conclusion]
1 / AC / Affirmative / "un-" negated	If a person is believing, then the person is unromantic	The person is unromantic	The person is believing
1 / AC / "not" negated / Affirmative	If a person is not believing, then the person is romantic	The person is romantic	The person is not believing
1 / AC / "not" negated / "not" negated	If a person is not believing, then the person is not romantic	The person is not romantic	The person is not believing
1 / AC / "not" negated / "un-" negated	If a person is not believing, then the person is unromantic	The person is unromantic	The person is not believing
1 / AC / "un-" negated / Affirmative	If a person is unbelieving, then the person is romantic	The person is romantic	The person is unbelieving
1 / AC / "un-" negated / "not" negated	If a person is unbelieving, then the person is not romantic	The person is not romantic	The person is unbelieving
1 / AC / "un-" negated / "un-" negated	If a person is unbelieving, then the person is unromantic	The person is unromantic	The person is unbelieving
1 / MT / Affirmative / Affirmative	If a person is believing, then the person is romantic	The person is not romantic	The person is not believing
1 / MT / Affirmative / "not" negated	If a person is believing, then the person is not romantic	The person is romantic	The person is not believing
1 / MT / Affirmative / "un-" negated	If a person is believing, then the person is unromantic	The person is romantic	The person is unbelieving
1 / MT / "not" negated / Affirmative	If a person is not believing, then the person is romantic	The person is not romantic	The person is believing

Scenario / Inference Type / Antecedent Polarity / Consequent Polarity	[conditional]	[second premise]	[conclusion]
1 / MT / "not" negated / "not" negated	If a person is not believing, then the person is not romantic	The person is romantic	The person is believing
1 / MT / "not" negated / "un-" negated	If a person is not believing, then the person is unromantic	The person is romantic	The person is believing
1 / MT / "un-" negated / Affirmative	If a person is unbelieving, then the person is romantic	The person is unromantic	The person is believing
1 / MT / "un-" negated / "not" negated	If a person is unbelieving, then the person is not romantic	The person is romantic	The person is believing
1 / MT / "un-" negated / "un-" negated	If a person is unbelieving, then the person is unromantic	The person is romantic	The person is believing
2 / MP / Affirmative / Affirmative	If a person is friendly, then the person is social	The person is friendly	The person is social
2 / MP / Affirmative / "not" negated	If a person is friendly, then the person is not social	The person is friendly	The person is not social
2 / MP / Affirmative / "un-" negated	If a person is friendly, then the person is unsocial	The person is friendly	The person is unsocial
2 / MP / "not" negated / Affirmative	If a person is not friendly, then the person is social	The person is not friendly	The person is social
2 / MP / "not" negated / "not" negated	If a person is not friendly, then the person is not social	The person is not friendly	The person is not social
2 / MP / "not" negated / "un-" negated	If a person is not friendly, then the person is unsocial	The person is not friendly	The person is unsocial

Scenario / Inference Type / Antecedent Polarity / Consequent Polarity	[conditional]	[second premise]	[conclusion]
2 / MP / "un-" negated / Affirmative	If a person is unfriendly, then the person is social	The person is unfriendly	The person is social
2 / MP / "un-" negated / "not" negated	If a person is unfriendly, then the person is not social	The person is unfriendly	The person is not social
2 / MP / "un-" negated / "un-" negated	If a person is unfriendly, then the person is unsocial	The person is unfriendly	The person is unsocial
2 / DA / Affirmative / Affirmative	If a person is friendly, then the person is social	The person is not friendly	The person is not social
2 / DA / Affirmative / "not" negated	If a person is friendly, then the person is not social	The person is not friendly	The person is social
2 / DA / Affirmative / "un-" negated	If a person is friendly, then the person is unsocial	The person is unfriendly	The person is social
2 / DA / "not" negated / Affirmative	If a person is not friendly, then the person is social	The person is friendly	The person is not social
2 / DA / "not" negated / "not" negated	If a person is not friendly, then the person is not social	The person is friendly	The person is social
2 / DA / "not" negated / "un-" negated	If a person is not friendly, then the person is unsocial	The person is friendly	The person is social
2 / DA / "un-" negated / Affirmative	If a person is unfriendly, then the person is social	The person is friendly	The person is unsocial
2 / DA / "un-" negated / "not" negated	If a person is unfriendly, then the person is not social	The person is friendly	The person is social
2 / DA / "un-" negated / "un-" negated	If a person is unfriendly, then the person is unsocial	The person is friendly	The person is social

Scenario / Inference Type / Antecedent Polarity / Consequent Polarity	[conditional]	[second premise]	[conclusion]
2 / AC / Affirmative / Affirmative	If a person is friendly, then the person is social	The person is social	The person is friendly
2 / AC / Affirmative / "not" negated	If a person is friendly, then the person is not social	The person is not social	The person is friendly
2 / AC / Affirmative / "un-" negated	If a person is friendly, then the person is unsocial	The person is unsocial	The person is friendly
2 / AC / "not" negated / Affirmative	If a person is not friendly, then the person is social	The person is social	The person is not friendly
2 / AC / "not" negated / "not" negated	If a person is not friendly, then the person is not social	The person is not social	The person is not friendly
2 / AC / "not" negated / "un-" negated	If a person is not friendly, then the person is unsocial	The person is unsocial	The person is not friendly
2 / AC / "un-" negated / Affirmative	If a person is unfriendly, then the person is social	The person is social	The person is unfriendly
2 / AC / "un-" negated / "not" negated	If a person is unfriendly, then the person is not social	The person is not social	The person is unfriendly
2 / AC / "un-" negated / "un-" negated	If a person is unfriendly, then the person is unsocial	The person is unsocial	The person is unfriendly
2 / MT / Affirmative / Affirmative	If a person is friendly, then the person is social	The person is not social	The person is not friendly
2 / MT / Affirmative / "not" negated	If a person is friendly, then the person is not social	The person is social	The person is not friendly
2 / MT / Affirmative / "un-" negated	If a person is friendly, then the person is unsocial	The person is social	The person is unfriendly



Scenario / Inference Type / Antecedent Polarity / Consequent Polarity	[conditional]	[second premise]	[conclusion]
2 / MT / "not" negated / Affirmative	If a person is not friendly, then the person is social	The person is not social	The person is friendly
2 / MT / "not" negated / "not" negated	If a person is not friendly, then the person is not social	The person is social	The person is friendly
2 / MT / "not" negated / "un-" negated	If a person is not friendly, then the person is unsocial	The person is social	The person is friendly
2 / MT / "un-" negated / Affirmative	If a person is unfriendly, then the person is social	The person is unsocial	The person is friendly
2 / MT / "un-" negated / "not" negated	If a person is unfriendly, then the person is not social	The person is social	The person is friendly
2 / MT / "un-" negated / "un-" negated	If a person is unfriendly, then the person is unsocial	The person is social	The person is friendly
3 / MP / Affirmative / Affirmative	If a person is grateful, then the person is happy	The person is grateful	The person is happy
3 / MP / Affirmative / "not" negated	If a person is grateful, then the person is not happy	The person is grateful	The person is not happy
3 / MP / Affirmative / "un-" negated	If a person is grateful, then the person is unhappy	The person is grateful	The person is unhappy
3 / MP / "not" negated / Affirmative	If a person is not grateful, then the person is happy	The person is not grateful	The person is happy
3 / MP / "not" negated / "not" negated	If a person is not grateful, then the person is not happy	The person is not grateful	The person is not happy
3 / MP / "not" negated / "un-" negated	If a person is not grateful, then the person is unhappy	The person is not grateful	The person is unhappy

Scenario / Inference Type / Antecedent Polarity / Consequent Polarity	[conditional]	[second premise]	[conclusion]
3 / MP / "un-" negated / Affirmative	If a person is ungrateful, then the person is happy	The person is ungrateful	The person is happy
3 / MP / "un-" negated / "not" negated	If a person is ungrateful, then the person is not happy	The person is ungrateful	The person is not happy
3 / MP / "un-" negated / "un-" negated	If a person is ungrateful, then the person is unhappy	The person is ungrateful	The person is unhappy
3 / DA / Affirmative / Affirmative	If a person is grateful, then the person is happy	The person is not grateful	The person is not happy
3 / DA / Affirmative / "not" negated	If a person is grateful, then the person is not happy	The person is not grateful	The person is happy
3 / DA / Affirmative / "un-" negated	If a person is grateful, then the person is unhappy	The person is ungrateful	The person is happy
3 / DA / "not" negated / Affirmative	If a person is not grateful, then the person is happy	The person is grateful	The person is not happy
3 / DA / "not" negated / "not" negated	If a person is not grateful, then the person is not happy	The person is grateful	The person is happy
3 / DA / "not" negated / "un-" negated	If a person is not grateful, then the person is unhappy	The person is grateful	The person is happy
3 / DA / "un-" negated / Affirmative	If a person is ungrateful, then the person is happy	The person is grateful	The person is unhappy
3 / DA / "un-" negated / "not" negated	If a person is ungrateful, then the person is not happy	The person is grateful	The person is happy
3 / DA / "un-" negated / "un-" negated	If a person is ungrateful, then the person is unhappy	The person is grateful	The person is happy

Scenario / Inference Type / Antecedent Polarity / Consequent Polarity	[conditional]	[second premise]	[conclusion]
3 / AC / Affirmative / Affirmative	If a person is grateful, then the person is happy	The person is happy	The person is grateful
3 / AC / Affirmative / "not" negated	If a person is grateful, then the person is not happy	The person is not happy	The person is grateful
3 / AC / Affirmative / "un-" negated	If a person is grateful, then the person is unhappy	The person is unhappy	The person is grateful
3 / AC / "not" negated / Affirmative	If a person is not grateful, then the person is happy	The person is happy	The person is not grateful
3 / AC / "not" negated / "not" negated	If a person is not grateful, then the person is not happy	The person is not happy	The person is not grateful
3 / AC / "not" negated / "un-" negated	If a person is not grateful, then the person is unhappy	The person is unhappy	The person is not grateful
3 / AC / "un-" negated / Affirmative	If a person is ungrateful, then the person is happy	The person is happy	The person is ungrateful
3 / AC / "un-" negated / "not" negated	If a person is ungrateful, then the person is not happy	The person is not happy	The person is ungrateful
3 / AC / "un-" negated / "un-" negated	If a person is ungrateful, then the person is unhappy	The person is unhappy	The person is ungrateful
3 / MT / Affirmative / Affirmative	If a person is grateful, then the person is happy	The person is not happy	The person is not grateful
3 / MT / Affirmative / "not" negated	If a person is grateful, then the person is not happy	The person is happy	The person is not grateful
3 / MT / Affirmative / "un-" negated	If a person is grateful, then the person is unhappy	The person is happy	The person is ungrateful

Scenario / Inference Type / Antecedent Polarity / Consequent Polarity	[conditional]	[second premise]	[conclusion]
3 / MT / "not" negated / Affirmative	If a person is not grateful, then the person is happy	The person is not happy	The person is grateful
3 / MT / "not" negated / "not" negated	If a person is not grateful, then the person is not happy	The person is happy	The person is grateful
3 / MT / "not" negated / "un-" negated	If a person is not grateful, then the person is unhappy	The person is happy	The person is grateful
3 / MT / "un-" negated / Affirmative	If a person is ungrateful, then the person is happy	The person is unhappy	The person is grateful
3 / MT / "un-" negated / "not" negated	If a person is ungrateful, then the person is not happy	The person is happy	The person is grateful
3 / MT / "un-" negated / "un-" negated	If a person is ungrateful, then the person is unhappy	The person is happy	The person is grateful
4 / MP / Affirmative / Affirmative	If a person is stylish, then the person is wise	The person is stylish	The person is wise
4 / MP / Affirmative / "not" negated	If a person is stylish, then the person is not wise	The person is stylish	The person is not wise
4 / MP / Affirmative / "un-" negated	If a person is stylish, then the person is unwise	The person is stylish	The person is unwise
4 / MP / "not" negated / Affirmative	If a person is not stylish, then the person is wise	The person is not stylish	The person is wise
4 / MP / "not" negated / "not" negated	If a person is not stylish, then the person is not wise	The person is not stylish	The person is not wise
4 / MP / "not" negated / "un-" negated	If a person is not stylish, then the person is unwise	The person is not stylish	The person is unwise

Scenario / Inference Type / Antecedent Polarity / Consequent Polarity	[conditional]	[second premise]	[conclusion]
4 / MP / "un-" negated / Affirmative	If a person is unstylish, then the person is wise	The person is unstylish	The person is wise
4 / MP / "un-" negated / "not" negated	If a person is unstylish, then the person is not wise	The person is unstylish	The person is not wise
4 / MP / "un-" negated / "un-" negated	If a person is unstylish, then the person is unwise	The person is unstylish	The person is unwise
4 / DA / Affirmative / Affirmative	If a person is stylish, then the person is wise	The person is not stylish	The person is not wise
4 / DA / Affirmative / "not" negated	If a person is stylish, then the person is not wise	The person is not stylish	The person is wise
4 / DA / Affirmative / "un-" negated	If a person is stylish, then the person is unwise	The person is unstylish	The person is wise
4 / DA / "not" negated / Affirmative	If a person is not stylish, then the person is wise	The person is stylish	The person is not wise
4 / DA / "not" negated / "not" negated	If a person is not stylish, then the person is not wise	The person is stylish	The person is wise
4 / DA / "not" negated / "un-" negated	If a person is not stylish, then the person is unwise	The person is stylish	The person is wise
4 / DA / "un-" negated / Affirmative	If a person is unstylish, then the person is wise	The person is stylish	The person is unwise
4 / DA / "un-" negated / "not" negated	If a person is unstylish, then the person is not wise	The person is stylish	The person is wise
4 / DA / "un-" negated / "un-" negated	If a person is unstylish, then the person is unwise	The person is stylish	The person is wise

Scenario / Inference Type / Antecedent Polarity / Consequent Polarity	[conditional]	[second premise]	[conclusion]
4 / AC / Affirmative / Affirmative	If a person is stylish, then the person is wise	The person is wise	The person is stylish
4 / AC / Affirmative / “not” negated	If a person is stylish, then the person is not wise	The person is not wise	The person is stylish
4 / AC / Affirmative / “un-” negated	If a person is stylish, then the person is unwise	The person is unwise	The person is stylish
4 / AC / “not” negated / Affirmative	If a person is not stylish, then the person is wise	The person is wise	The person is not stylish
4 / AC / “not” negated / “not” negated	If a person is not stylish, then the person is not wise	The person is not wise	The person is not stylish
4 / AC / “not” negated / “un-” negated	If a person is not stylish, then the person is unwise	The person is unwise	The person is not stylish
4 / AC / “un-” negated / Affirmative	If a person is unstylish, then the person is wise	The person is wise	The person is unstylish
4 / AC / “un-” negated / “not” negated	If a person is unstylish, then the person is not wise	The person is not wise	The person is unstylish
4 / AC / “un-” negated / “un-” negated	If a person is unstylish, then the person is unwise	The person is unwise	The person is unstylish
4 / MT / Affirmative / Affirmative	If a person is stylish, then the person is wise	The person is not wise	The person is not stylish
4 / MT / Affirmative / “not” negated	If a person is stylish, then the person is not wise	The person is wise	The person is not stylish
4 / MT / Affirmative / “un-” negated	If a person is stylish, then the person is unwise	The person is wise	The person is unstylish

Scenario / Inference Type / Antecedent Polarity / Consequent Polarity	[conditional]	[second premise]	[conclusion]
4 / MT / “not” negated / Affirmative	If a person is not stylish, then the person is wise	The person is not wise	The person is stylish
4 / MT / “not” negated / “not” negated	If a person is not stylish, then the person is not wise	The person is wise	The person is stylish
4 / MT / “not” negated / “un-” negated	If a person is not stylish, then the person is unwise	The person is wise	The person is stylish
4 / MT / “un-” negated / Affirmative	If a person is unstylish, then the person is wise	The person is unwise	The person is stylish
4 / MT / “un-” negated / “not” negated	If a person is unstylish, then the person is not wise	The person is wise	The person is stylish
4 / MT / “un-” negated / “un-” negated	If a person is unstylish, then the person is unwise	The person is wise	The person is stylish

### **A2.3.3. Task Instructions for Experiments 3 and 4, Part 2**

No further instructions were provided before the start of the plausibility tasks in experiment 3.

The following text was provided to participants in experiment 4 to explain the plausibility tasks:

*You have completed the first part of the survey.*

*The second part will consist of further questions which provide a scenario and a rule or statement and ask you a question about how plausible you think the rule or statement is given the scenario. You enter your response as a numerical answer between 0 and 100.*

*The scenarios given are each used in multiple questions. However, the rule and statement change in each question so you should pay attention to this.*

*All of the questions are asking for your view, there are no right or wrong answers and no questions are trying to catch you out.*

*You should try to complete each question as quickly as possible – go with your first reaction.*

*When you are ready to begin, click on 'Next' for the first question of the second (and final) part.*

#### **A2.3.4. Task Questions for Experiments 3 and 4, Part 2**

The plausibility questions in experiments 3 and 4 used the same 36 conditionals that were used for the inference tasks.

In experiment 3, participants saw the only the four conditionals which corresponded with those the participants saw in the inference tasks. The four questions seen were presented in random order.

In experiment 4, participants saw all 36 conditionals in random order. Intermixed with these were 12 further tasks based on statements of the antecedent propositions used in the conditionals in each of the .

Each task was presented one to a screen in the following form:

[scenario introduction]

*An observer looks at the 100 people and suggests the following rule:*

[conditional/proposition]

*For how many of the 100 people do you think the [rule/statement] will be true for (enter a number between 0 and 100)?*



Participants had to type a number between 0 and 100 and click on a ‘Next’ button to continue.

The placeholder [scenario introduction] were varied for each of the four scenario conditions. The same four sets of scenarios were used as in part 1 and so the text used to replace this placeholder is the same as given in table A2.6.

The placeholders [conditional/proposition] and [rule/statement] were varied between scenario, antecedent polarity and consequent polarity conditions for the conditionals and scenario and polarity conditions for the additional propositions used in experiment 4. The following table sets out the content used for this placeholder for each condition.

*Table A2.8 – Inference tasks in chapter 3, experiments 3 and 4, part 2*

Scenario / Antecedent Polarity / Consequent Polarity Or Scenario / Proposition Polarity	[conditional/proposition]	[rule/statement]
1 / Affirmative / Affirmative	If a person is believing, then the person is romantic	rule
1 / Affirmative / “not” negated	If a person is believing, then the person is not romantic	rule
1 / Affirmative / “un-” negated	If a person is believing, then the person is unromantic	rule
1 / “not” negated / Affirmative	If a person is not believing, then the person is romantic	rule
1 / “not” negated / “not” negated	If a person is not believing, then the person is not romantic	rule
1 / “not” negated / “un-” negated	If a person is not believing, then the person is unromantic	rule
1 / “un-” negated / Affirmative	If a person is unbelieving, then the person is romantic	rule
1 / “un-” negated / “not” negated	If a person is unbelieving, then the person is not romantic	rule

Scenario / Antecedent	[conditional/proposition]	[rule/statement]	
Polarity / Consequent			
Polarity			
Or			
Scenario / Proposition			
Polarity			
1 / "un-" negated / "un-" negated	If a person is unbelieving, then the person is unromantic	rule	
1 / Affirmative	a person is believing	statement	Expt. 4 only
1 / "not" negated	a person is not believing	statement	Expt. 4 only
1 / "un-" negated	a person is unbelieving	statement	Expt. 4 only
2 / Affirmative / Affirmative	If a person is friendly, then the person is social	rule	
2 / Affirmative / "not" negated	If a person is friendly, then the person is not social	rule	
2 / Affirmative / "un-" negated	If a person is friendly, then the person is unsocial	rule	
2 / "not" negated / Affirmative	If a person is not friendly, then the person is social	rule	
2 / "not" negated / "not" negated	If a person is not friendly, then the person is not social	rule	
2 / "not" negated / "un-" negated	If a person is not friendly, then the person is unsocial	rule	
2 / "un-" negated / Affirmative	If a person is unfriendly, then the person is social	rule	
2 / "un-" negated / "not" negated	If a person is unfriendly, then the person is not social	rule	
2 / "un-" negated / "un-" negated	If a person is unfriendly, then the person is unsocial	rule	
2 / Affirmative	A person is friendly	statement	Expt. 4 only
2 / "not" negated	A person is not friendly	statement	Expt. 4 only
2 / "un-" negated	A person is unfriendly	statement	Expt. 4 only

Scenario / Antecedent	[conditional/proposition]	[rule/statement]	
Polarity / Consequent			
Polarity			
Or			
Scenario / Proposition			
Polarity			
3 / Affirmative / Affirmative	If a person is grateful, then the person is happy	rule	
3 / Affirmative / "not" negated	If a person is grateful, then the person is not happy	rule	
3 / Affirmative / "un-" negated	If a person is grateful, then the person is unhappy	rule	
3 / "not" negated / Affirmative	If a person is not grateful, then the person is happy	rule	
3 / "not" negated / "not" negated	If a person is not grateful, then the person is not happy	rule	
3 / "not" negated / "un-" negated	If a person is not grateful, then the person is unhappy	rule	
3 / "un-" negated / Affirmative	If a person is ungrateful, then the person is happy	rule	
3 / "un-" negated / "not" negated	If a person is ungrateful, then the person is not happy	rule	
3 / "un-" negated / "un-" negated	If a person is ungrateful, then the person is unhappy	rule	
3 / Affirmative	A person is grateful	statement	Expt. 4 only
3 / "not" negated	A person is not grateful	statement	Expt. 4 only
3 / "un-" negated	A person is ungrateful	statement	Expt. 4 only
4 / Affirmative / Affirmative	If a person is stylish, then the person is wise	rule	
4 / Affirmative / "not" negated	If a person is stylish, then the person is not wise	rule	
4 / Affirmative / "un-" negated	If a person is stylish, then the person is unwise	rule	
4 / "not" negated / Affirmative	If a person is not stylish, then the person is wise	rule	

Scenario / Antecedent	[conditional/proposition]	[rule/statement]	
Polarity / Consequent			
Polarity			
Or			
Scenario / Proposition			
Polarity			
4 / "not" negated / "not" negated	If a person is not stylish, then the person is not wise	rule	
4 / "not" negated / "un-" negated	If a person is not stylish, then the person is unwise	rule	
4 / "un-" negated / Affirmative	If a person is unstylish, then the person is wise	rule	
4 / "un-" negated / "not" negated	If a person is unstylish, then the person is not wise	rule	
4 / "un-" negated / "un-" negated	If a person is unstylish, then the person is unwise	rule	
4 / Affirmative	A person is stylish	statement	Expt. 4 only
4 / "not" negated	A person is not stylish	statement	Expt. 4 only
4 / "un-" negated	A person is unstylish	statement	Expt. 4 only

## **Appendix 3. Materials Used in Chapter 4 Experiments**

This appendix contains the experimental materials used in each of the three experiments reported in chapter 4. These experiments tested Oaksford and Chater's (2007) prediction of MP inference suppression when implicit negation is used instead of "not" negation and participants are provided with appropriate frequency information. They then expanded this prediction to use implicit negation to vary responses to DA, AC and MT inferences.

Each experiment in this chapter presented a learning phase, where participants were expected to attend to a sequent of picture providing frequency information in relation to a scenario, a set of inference tasks based on the scenario and a verification task which asked participants about the frequencies observed at the learning stage.

Experiment 1 (reported in §4.2 with method in §4.2.3 and materials discussed in §4.2.3.3) presented tasks based on an 'animals' scenario. This 'animals' scenario was used again in experiment 2 (reported in §4.3 with method in §4.3.2 and materials discussed in §4.3.2.3) along with a 'vehicles' scenario. This 'vehicles' scenario was used again in experiment 3 (reported in §4.4 with method in §4.4.2 and materials discussed in §4.4.2.3) along with an abstract 'shapes' scenario. Because two of the scenarios were used in a very similar forms across two experiments each, this appendix presents materials by scenario rather than by experiment.

### **A3.1. 'Animals' Scenario**

The 'animals' scenario presented a set of different animals (cats, dogs and rabbits) in different colours (black, white and brown).

#### **A3.1.1. Task Instructions for Experiment 1**

The following text was provided to participants to participants at the beginning of experiment 1 on a single screen:

*This study will ask several sets of questions based on the following scenario:*

*A vet sees a range of animals, mostly dogs, cats and rabbits. She notices that people in her village favour animals of certain colours. She wants to*

*come up with a rule linking the type of animal to its colour. She wants this to correctly link animal type and colour as often as possible.*

*There are four sets of questions:*

1. *Here you will be provided with some information about the percentages of different animals of different colours that the vet sees. You will then be provided with a rule which the vet could propose. You will then be given a partial description of an animal, e.g. its colour (like grey) or species (like horse). Finally, you will be asked to indicate what you could conclude about this animal. For example, you may be told that the animal is a horse and asked whether you believe it is grey or not. You will also be asked how confident you are in that belief.*

2. *Here you will be presented with pictures of a set of animals that the vet has seen. You will then be asked you to identify their colour and species.*

3. *Here you will be see a set of questions that are very similar to the first set. Each question provides a rule which the vet could propose and a partial description of the next animal that the vet will see. You will be asked what you would conclude about this animal and your confidence in that conclusion.*

4. *Here you will be asked to indicate the proportion of different coloured animals that you have seen.*

*Each question is asking for your view and no questions are trying to catch you out.*

*You should try to complete each question as quickly as possible – go with your first reaction.*

*When you are ready to continue to the first set of questions, click on 'Next'.*

After these initial instructions, the following instructions were provided on screen before the first set of inference questions:

*This first set of questions are all based on the following information:*

*The vet's receptionist tallies the animals that the vet sees over a week and tells the vet that the animals she saw were made up of the following types:*

- *Brown cats - 4%*
- *White dogs - 2%*
- *Black cats - 30%*
- *Brown rabbits - 22%*
- *White cats - 6%*
- *Black dogs - 10%*
- *White rabbits - 22%*
- *Brown dogs - 4%*

*Based on this information, the vet proposes the following rule:*

- *If it is not a cat, then it is not black.*

*The following questions provide information about animals that the vet is due to see as she is considering this rule. You should indicate what you think the vet should conclude in each case and how confident you would be that the conclusion selected is correct*

*When you are ready to begin, click on 'Next' for the first question.*

After the participant had completed the inference questions for the first time, they saw the following instructions before the learning phase:

*Thank you for completing the first set of questions.*

*For the next set of questions:*

*The vet observes the colour and type of the next 50 animals that she sees.*

*The following questions each provide a picture of one of the fifty animals that the vet observes and asks you to identify the animal and colour. Please select the options that you think best describes the animal even if you don't think the description is perfect.*

*When you are ready to begin, click on 'Next' for the first question in this set.*

After the learning phase, participants saw a screen with the following instructions before seeing the inference questions for the second time:

*Thank you for completing the second set of questions.*

*Having seen these animals, the vet again considers the following rule:*

- *If it is not a cat, then it is not black.*

*The following questions provide information about animals that the vet is due to see as she is considering this rule. You should indicate what you think the vet should conclude in each case and how confident you would be that the conclusion selected is correct*

*When you are ready to begin, click on 'Next' for the next set of questions.*

After the second presentation of the inference questions, participants were presented with the verification task on a final page.

### **A3.1.2. Task Instructions for Experiment 2**

The following text was provided to participants to participants at the beginning of experiment 2 on a single screen:

*This study will ask several sets of questions based two scenarios. One scenario is based around a vet considering the animals she sees and the other scenario is based around a police traffic officer considering the vehicles she sees.*

*For each scenario, there are three sets of questions:*



1. *First you will be presented with pictures of a set of animals that the vet has seen or vehicles that the police traffic officer has seen. You will then be asked you to identify their colour and type.*
2. *Then you will be see a set of questions which provide a rule linking the colour and type of the animals or vehicles observed and a partial description of the next animal or vehicle seen. You will be asked what you would conclude about this animal or vehicle and your confidence in that conclusion.*
3. *Here you will be asked to indicate the proportion of different coloured animals or vehicles of different types that you have seen.*

*Once you have completed each set of questions for one scenario, you will be asked to complete similar sets of questions for the other scenario.*

*Each question is asking for your view and no questions are trying to catch you out.*

*You should try to complete each question as quickly as possible – go with your first reaction.*

*When you are ready to continue to the first set of questions, click on 'Next'.*

After this, the participant completed questions for either the ‘animal’ or ‘vehicle’ scenario. The scenarios were randomly ordered. Once they had completed one scenario they completed the other. At the start of the ‘animal’ scenario, the following instructions were provided:

*For this set of questions, please consider the following scenario:*

*A vet sees a range of animals, mostly dogs, cats and rabbits. She notices that people in her village favour animals of certain colours. She wants to come up*

*with a rule linking the type of animal to its colour. She wants this rule to correctly link animal type and colour as often as possible.*

*The vet observes the colour and type of the next 30 animals that she sees.*

*The following questions each provide a picture of one of the thirty animals that the vet observes and asks you to identify the animal and colour. Please select the options that you think best describes the animal even if you don't think the description is perfect.*

*When you are ready to begin, click on 'Next' for the first question in this set.*

After the learning phase, participants were provided with the following instructions before the inference questions:

*Thank you for completing the first set of questions with this scenario.*

*Having seen these animals, the vet considers the following rule:*

- *If it is not a cat, then it is not black.*

*The following questions provide information about animals that the vet is due to see as she is considering this rule. You should indicate what you think the vet should conclude in each case and how confident you would be that the conclusion selected is correct*

*When you are ready to begin, click on 'Next' for the next set of questions.*

After this, participants completed the verification task before moving onto the other scenario or completing the experiment.

### **A3.1.3. Learning Phase**

The learning phase consisted of pages presented in random order. Each page had a picture of an animal with two questions below it. Participants had to pick an option to

respond to each question and click ‘Next’ to move on. The questions (with available response options in bullet points – which were presented in random order on each page) were:

*What type of animal is this?*



- *Cat*
- *Dog*
- *Rabbit*






*Which colour best describes this animal?*






- *Black*
- *White*
- *Brown*





The following table shows the pictures of animals used in this phase. All pictures were used in the experiment 1 learning phase. The table indicates whether each picture used in the MP manipulation condition, the AC manipulation condition, both or neither in experiment 2.






*Table A3.1 – Learning task stimuli for ‘animals’ scenario*






Picture Label	Picture	Experiment 2 Conditions Used
Black Cat 1		Both
Black Cat 2		Both

Picture Label	Picture	Experiment 2 Conditions Used
Black Cat 3		Both
Black Cat 4		Both
Black Cat 5		Both
Black Cat 6		Both
Black Cat 7		Both






Picture Label	Picture	Experiment 2 Conditions Used
Black Cat 8		Both
Black Cat 9		Neither
Black Cat 10		Neither
Black Cat 11		Neither
Black Cat 12		Neither






Picture Label	Picture	Experiment 2 Conditions Used
Black Cat 13		Neither
Black Cat 14		Neither
Black Cat 15		Neither
Black Dog 1		MP manipulation
Black Dog 2		MP manipulation






Picture Label	Picture	Experiment 2 Conditions Used
Black Dog 3		Neither
Black Dog 4		Neither
Black Dog 5		Neither
Brown Cat 1		Neither
Brown Cat 2		Neither






Picture Label	Picture	Experiment 2 Conditions Used
Brown Dog 1		AC manipulation
Brown Dog 2		AC manipulation
Brown Dog 3		AC manipulation
Brown Dog 4		AC manipulation
Brown Dog 5		AC manipulation








Picture Label	Picture	Experiment 2 Conditions Used
Brown Dog 6		AC manipulation
Brown Dog 7		AC manipulation
Brown Dog 8		AC manipulation
Brown Dog 9		AC manipulation
Brown Dog 10		AC manipulation


Picture Label	Picture	Experiment 2 Conditions Used
Brown Rabbit 1		Both
Brown Rabbit 2		Both
Brown Rabbit 3		Both
Brown Rabbit 4		Both
Brown Rabbit 5		Both

Picture Label	Picture	Experiment 2 Conditions Used
Brown Rabbit 6		Both
Brown Rabbit 7		Both
Brown Rabbit 8		Both
Brown Rabbit 9		Both
Brown Rabbit 10		Both

Picture Label	Picture	Experiment 2 Conditions Used
Brown Rabbit 11		Neither
White Cat 1		AC manipulation
White Cat 2		AC manipulation
White Cat 3		Neither
White Dog 1		Neither

Picture Label	Picture	Experiment 2 Conditions Used
White Rabbit 1		MP manipulation
White Rabbit 2		MP manipulation
White Rabbit 3		MP manipulation
White Rabbit 4		MP manipulation
White Rabbit 5		MP manipulation

Picture Label	Picture	Experiment 2 Conditions Used
White Rabbit 6		MP manipulation
White Rabbit 7		MP manipulation
White Rabbit 8		MP manipulation
White Rabbit 9		MP manipulation
White Rabbit 10		MP manipulation

Picture Label	Picture	Experiment 2 Conditions Used
White Rabbit		Neither
11		

#### **A3.1.4. Inference Questions**

Six inference tasks were prepared for the animal scenario. In experiment 1, these tasks were presented twice: before and after the learning phase. In experiment 2, they were only presented after the learning phase. In each case they were presented one to a page in random order. Each inference task page said the following:

*The vet is considering the following rule about the animals that she sees:*

- *If it is not a cat, then it is not black.*

*The vet is told that the next animal she will see is:*

- [premise]

*Please select the option below that best describes what she should conclude about the next animal.*

[response options]

*How confident are you that this conclusion will be correct? Please use the slider below to indicate your confidence.*

At the bottom of the page was a slider for participants to indicate their confidence. Participants had to select one of the response options, select a point on the slider and click 'Next' to move on.

The placeholders [premise] and [response options] were varied between inference and negation type conditions. The following table sets out the content used for this placeholder for each condition.

*Table A3.2 – Inference task stimuli for ‘animals’ scenario*

Condition	Premise	Response Options
MP (not)	not a cat.	<ul style="list-style-type: none"> <li>• That the animal is not black</li> <li>• That the animal is black</li> </ul>
MP (contrast class)	a dog.	<ul style="list-style-type: none"> <li>• That the animal is not black</li> <li>• That the animal is black</li> </ul>
DA	a cat	<ul style="list-style-type: none"> <li>• That the animal is not black</li> <li>• That the animal is black</li> </ul>
AC (not)	not black.	<ul style="list-style-type: none"> <li>• That the animal is not a cat</li> <li>• That the animal is a cat</li> </ul>
AC (contrast class)	white.	<ul style="list-style-type: none"> <li>• That the animal is not a cat</li> <li>• That the animal is a cat</li> </ul>
MT	black.	<ul style="list-style-type: none"> <li>• That the animal is not a cat</li> <li>• That the animal is a cat</li> </ul>

#### **A3.1.5. Verification Task**

At the end of experiment 1 and after the ‘animal’ scenario inference questions in experiment 2, a verification task was presented. The task was presented on a single page. The verification task page said the following:

*This page contains the final question. [In experiment 1 only]*

*This page contains the final question for this scenario. [In experiment 2 only]*

*For these questions, imagine the vet sees another one hundred animals. Please indicate how many of these animals you think are likely to fall into each of the categories below?*

*Your answers should add up to 100.*



The page then had a grid with types of animal (cat, dog, rabbit) listed in random order along the top and colours (black, white, brown) listed in random order down the side. Each of the nine cells had a box into which participants could enter a number.

Participants had to fill in the grid and press 'Next' to move on. If the numbers provided did not sum to 100 the page remained up with a message at the top to 'Please make sure all responses in the table add up to 100.'

### **A3.2. 'Vehicles' Scenario**

The 'vehicles' scenario presented a set of different vehicles (cars, vans and motorbikes) in different colours (red, blue and white). This scenario was used in experiments 2 and 3.

#### **A3.2.1. Task Instructions**

The following text was provided to participants to participants at the beginning of experiment 2 on a single screen:

*This study will ask several sets of questions based two scenarios. One scenario is based around a vet considering the animals she sees and the other scenario is based around a police traffic officer considering the vehicles she sees.*

*For each scenario, there are three sets of questions:*

- 1. First you will be presented with pictures of a set of animals that the vet has seen or vehicles that the police traffic officer has seen. You will then be asked you to identify their colour and type.*
- 2. Then you will be see a set of questions which provide a rule linking the colour and type of the animals or vehicles observed and a partial description of the next animal or vehicle seen. You will be asked what you would conclude about this animal or vehicle and your confidence in that conclusion.*
- 3. Here you will be asked to indicate the proportion of different coloured animals or vehicles of different types that you have seen.*

*Once you have completed each set of questions for one scenario, you will be asked to complete similar sets of questions for the other scenario.*

*Each question is asking for your view and no questions are trying to catch you out.*

*You should try to complete each question as quickly as possible – go with your first reaction.*

*When you are ready to continue to the first set of questions, click on 'Next'.*

The following text was provided at the start of experiment 3 on a single screen:

*This study will ask several sets of questions based two scenarios. One scenario is based around a quality control manager considering the products that she sees produced and the other scenario is based around a police traffic officer considering the vehicles she sees.*

*For each scenario, there are three sets of questions:*

- 1. First you will be presented with pictures of a set of products that the quality control manager has seen or vehicles that the police traffic officer has seen. You will then be asked you to identify their colour and type.*
- 2. Then you will be see a set of questions which provide a rule linking the colour and type of the products or vehicles observed and a partial description of the next product or vehicle seen. You will be asked what you would conclude about this animal or vehicle and your confidence in that conclusion.*
- 3. Here you will be asked to indicate the proportion of different coloured products or vehicles of different types that you have seen.*

*Once you have completed each set of questions for one scenario, you will be asked to complete similar sets of questions for the other scenario.*

*Each question is asking for your view and no questions are trying to catch you out.*

*You should try to complete each question as quickly as possible – go with your first reaction.*

*When you are ready to continue to the first set of questions, click on 'Next'.*

After this, the participant completed questions for either the ‘vehicle’ or alternative scenario. The scenarios were randomly ordered. Once they had completed one scenario they completed the other. At the start of the ‘vehicle’ scenario, the following instructions were provided:

*For this set of questions, please consider the following scenario:*

*A police traffic officer sees a range of vehicles, mostly cars, vans and motorbikes. She notices that drivers on the road she is monitoring favour vehicles of certain colours. She wants to come up with a rule linking the type of vehicle to its colour. She wants this rule to correctly link vehicle type and colour as often as possible.*

*The police traffic officers observes the colour and type of the next 30 vehicles that she sees.*

*The following questions each provide a picture of one of the thirty vehicles that the police traffic officer observes and asks you to identify the vehicle and colour. Please select the options that you think best describes the vehicle even if you don't think the description is perfect.*

*When you are ready to begin, click on 'Next' for the first question in this set.*

After the learning phase, participants were provided with the following instructions before the inference questions:

*Thank you for completing the first set of questions with this scenario.*

*Having seen these vehicles, the police traffic officer considers the following rule:*

- *If it is not white, then it is not a van. [in experiment 2 only]*
- *If it is white, then it is a van. [in experiment 3 only]*

*The following questions provide information about vehicles that the police traffic officer is due to see as she is considering this rule. You should indicate what you think the police traffic officer should conclude in each case and how confident you would be that the conclusion selected is correct*

*When you are ready to begin, click on 'Next' for the next set of questions.*

After this, participants completed the verification task before moving onto the other scenario or completing the experiment.

### **A3.2.2. Learning Phase**

The learning phase consisted of pages presented in random order. Each page had a picture of an vehicle with two questions below it. Participants had to pick an option to respond to each question and click 'Next' to move on. The questions (with available response options in bullet points – which were presented in random order on each page) were:

*What type of vehicle is this?*



- *Van*
- *Motorbike*
- *Car*

*Which colour best describes this vehicle?*






- *White*
- *Blue*
- *Red*






The following table shows the pictures of vehicles used in this phase. The table indicates whether each picture used in the MP manipulation condition, the AC manipulation condition or both experiment 2. The table also indicates whether each picture used in the DA manipulation condition, the MT manipulation condition or both in experiment 3.

*Table A3.3 – Learning task stimuli for ‘vehicles’ scenario*






Picture Label	Picture	Experiment 2 Conditions Used	Experiment 3 Conditions Used
Blue Car 1		AC manipulation	MT manipulation
Blue Car 2		AC manipulation	MT manipulation






Picture Label	Picture	Experiment 2 Conditions Used	Experiment 3 Conditions Used
Blue Car 3		AC manipulation	MT manipulation
Blue Car 4		AC manipulation	MT manipulation
Blue Car 5		AC manipulation	MT manipulation
Blue Car 6		AC manipulation	MT manipulation
Blue Car 7		AC manipulation	MT manipulation






Picture Label	Picture	Experiment 2 Conditions Used	Experiment 3 Conditions Used
Blue Car 8		AC manipulation	MT manipulation
Blue Car 9		AC manipulation	MT manipulation
Blue Car 10		AC manipulation	MT manipulation
Blue Van 1		MP Manipulation	DA manipulation
Blue Van 2		MP Manipulation	DA manipulation






Picture Label	Picture	Experiment 2 Conditions Used	Experiment 3 Conditions Used
Red Car 1		Both	Both
Red Car 2		Both	Both
Red Car 3		Both	Both
Red Car 4		Both	Both
Red Car 5			







Picture Label	Picture	Experiment 2 Conditions Used	Experiment 3 Conditions Used
Red Car 6		Both	Both
Red Car 7		Both	Both
Red Car 8		Both	Both
Red Car 9		Both	Both
Red Car 10		Both	Both

Picture Label	Picture	Experiment 2 Conditions Used	Experiment 3 Conditions Used
Red Motorbike 1		MP manipulation	DA manipulation
Red Motorbike 2		MP manipulation	DA manipulation
Red Motorbike 3		MP manipulation	DA manipulation
Red Motorbike 4		MP manipulation	DA manipulation
Red Motorbike 5		MP manipulation	DA manipulation

Picture Label	Picture	Experiment 2 Conditions Used	Experiment 3 Conditions Used
Red Motorbike 6		MP manipulation	DA manipulation
Red Motorbike 7		MP manipulation	DA manipulation
Red Motorbike 8		MP manipulation	DA manipulation
Red Motorbike 9		MP manipulation	DA manipulation
Red Motorbike 10		MP manipulation	DA manipulation

Picture Label	Picture	Experiment 2 Conditions Used	Experiment 3 Conditions Used
White Motorbike 1		AC manipulation	MT manipulation
White Motorbike 2		AC manipulation	MT manipulation
White Van 1		Both	Both
White Van 2		Both	Both
White Van 3		Both	Both

Picture Label	Picture	Experiment 2 Conditions Used	Experiment 3 Conditions Used
White Van 4		Both	Both
White Van 5		Both	Both
White Van 6		Both	Both
White Van 7		Both	Both

Picture Label	Picture	Experiment 2	Experiment 3
		Conditions	Conditions
		Used	Used
White Van 8		Both	Both



### **A3.2.3. Inference Questions**

Six inference tasks were prepared for each of experiments 2 and 3 using the ‘vehicle’ scenario. In both of these experiments, inference questions were presented after the learning phase. In each case they were presented one to a page in random order. Each inference task page using the ‘vehicle’ said the following:

*The police traffic officer is considering the following rule about the vehicles that she sees:*

- *If it is not white, then it is not a van. [in experiment 2 only]*
- *If it is white, then it is a van. [in experiment 3 only]*

*The police traffic officer is told that the next vehicle she will see is:*

- [premise]

*Please select the option below that best describes what she should conclude about the next vehicle.*

[response options]

*How confident are you that this conclusion will be correct? Please use the slider below to indicate your confidence.*

At the bottom of the page was a slider for participants to indicate their confidence. Participants had to select one of the response options, select a point on the slider and click ‘Next’ to move on.

The placeholders [premise] and [response options] were varied between inference and negation type conditions differently for experiments 2 and 3. The following table sets out the content used for this placeholder for each condition.

*Table A3.4 – Inference task stimuli for ‘vehicles’ scenario*

Experiment	Condition	Premise	Response Options
2	MP (not)	not white.	<ul style="list-style-type: none"> <li>• That the vehicle is not a van</li> <li>• That the vehicle is a van</li> </ul>
2	MP (contrast class)	blue.	<ul style="list-style-type: none"> <li>• That the vehicle is not a van</li> <li>• That the vehicle is a van</li> </ul>
2	DA	white.	<ul style="list-style-type: none"> <li>• That the vehicle is not a van</li> <li>• That the vehicle is a van</li> </ul>
2	AC (not)	not a van.	<ul style="list-style-type: none"> <li>• That the vehicle is not white</li> <li>• That the vehicle is white</li> </ul>
2	AC (contrast class)	a motorbike.	<ul style="list-style-type: none"> <li>• That the vehicle is not white</li> <li>• That the vehicle is white</li> </ul>
2	MT	a van.	<ul style="list-style-type: none"> <li>• That the vehicle is not white</li> <li>• That the vehicle is white</li> </ul>
3	MP	white.	<ul style="list-style-type: none"> <li>• That the vehicle is not a van</li> <li>• That the vehicle is a van</li> </ul>
3	DA (not)	not white.	<ul style="list-style-type: none"> <li>• That the vehicle is not a van</li> <li>• That the vehicle is a van</li> </ul>
3	DA (contrast class)	blue.	<ul style="list-style-type: none"> <li>• That the vehicle is not a van</li> <li>• That the vehicle is a van</li> </ul>
3	AC	a van.	<ul style="list-style-type: none"> <li>• That the vehicle is not white</li> <li>• That the vehicle is white</li> </ul>
3	MT (not)	not a van.	<ul style="list-style-type: none"> <li>• That the vehicle is not white</li> <li>• That the vehicle is white</li> </ul>
3	MT (contrast class)	a motorbike.	<ul style="list-style-type: none"> <li>• That the vehicle is not white</li> <li>• That the vehicle is white</li> </ul>

#### **A3.2.4. Verification Task**

After the ‘vehicle’ scenario inference questions in experiments 2 and 3, a verification task was presented. In each case this were presented on a single page. The verification task page said the following:

*This page contains the final question for this scenario.*

*For these questions, imagine the police traffic officer sees another one hundred vehicles. Please indicate how many of these vehicles you think are likely to fall into each of the categories below?*

*Your answers should add up to 100.*

The page then had a grid with types of vehicle (van, motorbike, car) listed in random order along the top and colours (blue, white, red) listed in random order down the side. Each of the nine cells had a box into which participants could enter a number.

Participants had to fill in the grid and press ‘Next’ to move on. If the numbers provided did not sum to 100 the page remained up with a message at the top to ‘Please make sure all responses in the table add up to 100.’

#### **A3.3. ‘Shapes’ Scenario**

The ‘shapes’ scenario presented a set of different vehicles (circle, square and triangle) in different colours (red, blue and yellow). This scenario was used in experiment 3.

##### **A3.3.1. Task Instructions**

The following text was provided to participants to participants at the beginning of experiment 3 on a single screen:

*This study will ask several sets of questions based two scenarios. One scenario is based around a quality control manager considering the products that she sees produced and the other scenario is based around a police traffic officer considering the vehicles she sees.*



*For each scenario, there are three sets of questions:*

- 1. First you will be presented with pictures of a set of products that the quality control manager has seen or vehicles that the police traffic officer has seen. You will then be asked you to identify their colour and type.*
- 2. Then you will be see a set of questions which provide a rule linking the colour and type of the products or vehicles observed and a partial description of the next product or vehicle seen. You will be asked what you would conclude about this animal or vehicle and your confidence in that conclusion.*
- 3. Here you will be asked to indicate the proportion of different coloured products or vehicles of different types that you have seen.*

*Once you have completed each set of questions for one scenario, you will be asked to complete similar sets of questions for the other scenario.*

*Each question is asking for your view and no questions are trying to catch you out.*

*You should try to complete each question as quickly as possible – go with your first reaction.*

*When you are ready to continue to the first set of questions, click on 'Next'.*

After this, the participant completed questions for either the ‘shapes’ or ‘vehicles’ scenario. The scenarios were randomly ordered. Once they had completed one scenario they completed the other. At the start of the ‘shapes’ scenario, the following instructions were provided:

*For this set of questions, please consider the following scenario:*

*A quality control manager works in a factory and is responsible for checking the output of a machine which produces cards with different shapes on them, mostly circles, squares and triangles. She is new to this machine but knows that it randomly produces shapes with different colours but the machine has certain rules dictating which type of shape can be take which colours. She wants to figure out one of the rules that the machine uses to link the type of shape with its colour.*

*The quality control manager observes the colour and type of the shapes on the next 30 cards that she sees.*

*The following questions each provide a picture of one of the thirty cards that the quality control manager observes and asks you to identify the shape and colour. Please select the options that you think best describes the shape even if you don't think the description is perfect.*

*When you are ready to begin, click on 'Next' for the first question in this set.*

After the learning phase, participants were provided with the following instructions before the inference questions:

*Thank you for completing the first set of questions with this scenario.*

*Having seen these cards, the quality control manager considers the following rule:*

- *If it is red, then it is a circle.*

*The following questions provide information about cards that the quality control manager is due to see as she is considering this rule. You should indicate what you think the quality control manager should conclude in each case and how confident you would be that the conclusion selected is correct*

*When you are ready to begin, click on 'Next' for the next set of questions.*

After this, participants completed the verification task before moving onto the other scenario or completing the experiment.

### **A3.3.2. Learning Phase**

The learning phase consisted of pages presented in random order. Each page had a picture of a card with a shape on it with two questions below it. Participants had to pick an option to respond to each question and click ‘Next’ to move on. The questions (with available response options in bullet points – which were presented in random order on each page) were:

*What type of shape is this?*


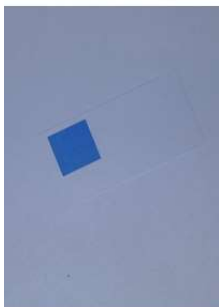
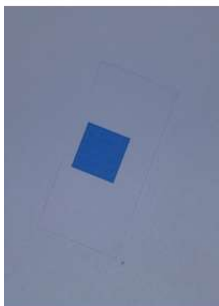
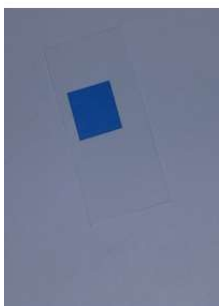
- *Circle*
- *Square*
- *Triangle*


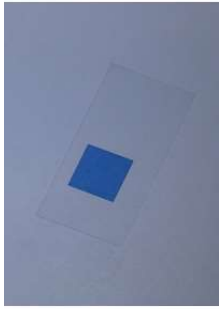
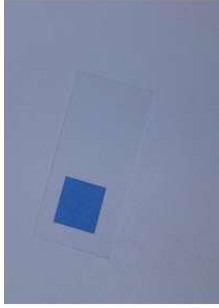


*Which colour best describes this shape?*






- *Yellow*
- *Blue*
- *Red*




The following table shows the pictures of cards with shapes on used in this phase. The table indicates whether each picture used in the DA manipulation condition, the MT manipulation condition or both in experiment 3.






*Table A3.5 – Learning task stimuli for ‘shapes’ scenario*

Picture Label	Picture	Experiment 3 Conditions Used
Blue Square 1		
Blue Square 2		
Blue Square 3		DA manipulation
Blue Square 4		DA manipulation



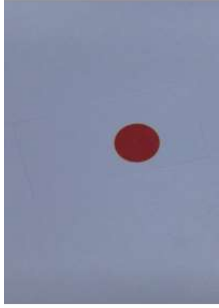


Picture Label	Picture	Experiment 3 Conditions Used
Blue Square 5		DA manipulation
Blue Square 6		DA manipulation
Blue Square 7		DA manipulation
Blue Square 8		DA manipulation
Blue Square 9		DA manipulation


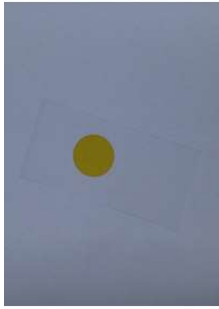



Picture Label	Picture	Experiment 3 Conditions Used
Blue Square 10		DA manipulation
Blue Triangle 1		Both
Blue Triangle 2		Both
Blue Triangle 3		Both
Blue Triangle 4		Both






Picture Label	Picture	Experiment 3 Conditions Used
Blue Triangle 5		Both
Blue Triangle 6		Both
Blue Triangle 7		Both
Blue Triangle 8		Both
Blue Triangle 9		Both




Picture Label	Picture	Experiment 3 Conditions Used
Blue Triangle 10		Both
Red Circle 1		Both
Red Circle 3		Both
Red Circle 3		Both
Red Circle 4		Both



Picture Label	Picture	Experiment 3 Conditions Used
Red Circle 5		Both
Red Circle 6		Both
Red Circle 7		Both
Red Circle 8		Both
Red Square 1		MT manipulation

Picture Label	Picture	Experiment 3 Conditions Used
Red Square 2		MT manipulation
Yellow Circle 1		DA manipulation
Yellow Circle 2		DA manipulation
Yellow Triangle 1		MT manipulation
Yellow Triangle 2		MT manipulation

Picture Label	Picture	Experiment 3 Conditions Used
Yellow Triangle 3		MT manipulation
Yellow Triangle 4		MT manipulation
Yellow Triangle 5		MT manipulation
Yellow Triangle 6		MT manipulation
Yellow Triangle 7		MT manipulation

Picture Label	Picture	Experiment 3 Conditions Used
Yellow Triangle 8		MT manipulation
Yellow Triangle 9		MT manipulation
Yellow Triangle 10		MT manipulation

### **A3.3.3. Inference Questions**

Six inference tasks were prepared experiment 3 using the ‘shapes’ scenario. These were presented after the learning phase. They were presented one to a page in random order. Each inference task page using the ‘shapes’ scenario said the following:

*The quality control manager is considering the following rule about the cards that she sees:*

- *If it is red, then it is a circle.*

*A system on the machine indicates that the shape on the next card is:*

- [premise]

*Please select the option below that best describes what she should conclude about the next shape.*

[response options]

*How confident are you that this conclusion will be correct? Please use the slider below to indicate your confidence.*

At the bottom of the page was a slider for participants to indicate their confidence. Participants had to select one of the response options, select a point on the slider and click ‘Next’ to move on.

The placeholders [premise] and [response options] were varied between inference and negation type conditions for experiment 3. The following table sets out the content used for these placeholders for each condition.

*Table A3.6 – Inference task stimuli for ‘shapes’ scenario*

Experiment	Condition	Premise	Response Options
3	MP	red.	<ul style="list-style-type: none"> <li>• That the shape is a circle</li> <li>• That the shape is not a circle</li> </ul>
3	DA (not)	not red.	<ul style="list-style-type: none"> <li>• That the shape is a circle</li> <li>• That the shape is not a circle</li> </ul>
3	DA (contrast class)	yellow.	<ul style="list-style-type: none"> <li>• That the shape is a circle</li> <li>• That the shape is not a circle</li> </ul>
3	AC	a circle.	<ul style="list-style-type: none"> <li>• That the shape is red</li> <li>• That the shape is not red</li> </ul>
3	MT (not)	not a circle.	<ul style="list-style-type: none"> <li>• That the shape is red</li> <li>• That the shape is not red</li> </ul>
3	MT (contrast class)	a square.	<ul style="list-style-type: none"> <li>• That the shape is red</li> <li>• That the shape is not red</li> </ul>

#### **A3.3.4. Verification Task**

After the ‘shapes’ scenario inference questions in experiment 3, a verification task was presented. This was presented on a single page. The verification task page said the following:

*This page contains the final question for this scenario.*

*For these questions, imagine the quality control manager sees another one hundred cards. Please indicate how many of the shapes of each type on the cards you think are likely to fall into each of the categories below?*

*Your answers should add up to 100.*

The page then had a grid with types of shape (circle, square, triangle) listed in random order along the top and colours (blue, yellow, red) listed in random order down the side. Each of the nine cells had a box into which participants could enter a number.

Participants had to fill in the grid and press ‘Next’ to move on. If the numbers provided did not sum to 100 the page remained up with a message at the top to ‘Please make sure all responses in the table add up to 100.’